



Three Essays in Political Economy and Public Finance

Citation

Troiano, Ugo A. 2013. Three Essays in Political Economy and Public Finance. Doctoral dissertation, Harvard University.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11125981>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Three Essays in Political Economy and Public Finance

A dissertation presented

by

Ugo A. Troiano

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

April 2013

© 2013 Ugo A. Troiano

All rights reserved.

Dissertation Advisors:
Professor Alberto Alesina
Professor Raj Chetty

Author:
Ugo A. Troiano

Three Essays in Political Economy and Public Finance

Abstract

Chapter 1 evaluates the effect of relaxing fiscal rules on policy outcomes applying a quasi-experimental research design. We implement a “difference-in-discontinuities” design by combining the before/after with the discontinuous policy variation generated by the implementation of the Domestic Stability Pact on Italian municipalities between 1999 and 2004. Our estimates show that relaxing fiscal rules triggers a substantial deficit bias, captured by a shift from a balanced budget to a deficit that amounts to 2 percent of the total budget. The deficit comes primarily from reduced revenues as unconstrained municipalities have lower real estate and income tax rates. The impact is larger if the mayor can run for reelection, the number of political parties seated in the city council is higher, voters are older, the performance of the mayor in providing public good is lower, and cities are characterized by historical deficit, consistently with models on the political economy of fiscal adjustment. Chapter 2 studies the electoral response to the Ghost Buildings program, a nationwide anti tax evasion policy in Italy, which used innovative monitoring technologies to target buildings hidden from tax authorities. The difference-in-differences identification strategy exploits both variation across towns in the ex ante program scope to increase enforcement as well as administrative data on actual building registrations. After the policy, local incumbents experience an increase in their reelection likelihood. These political returns are higher in areas with higher speed of public good provision and with lower tax evasion tolerance, implying complementarity among enforcement policies, government efficiency, and the underlying tax culture. Chapter 3 examines reasons for cross-country variation

in maternity leave provision. We show that the less tolerant a society is of gender-based discrimination, the longer the maternity leave it will optimally mandate. We collected new data on the number of gender-differentiated personal pronouns across languages to capture societies' attitudes toward gender-based discrimination. We first confirm, using within-country language variation, that our linguistic measure is correlated with attitudes toward gender-based discrimination. Then, using cross-country data on length of maternity leave we find a strong correlation between our measure of attitudes and the length of maternity leave.

Contents

Abstract	iii
Acknowledgments	x
1 Policy Responses to Fiscal Restraints: A Difference-in-Discontinuities Design	1
1.1 Introduction	1
1.2 Related literature	6
1.3 Institutional framework	7
1.4 Difference-in-discontinuities design	11
1.4.1 Setup	11
1.4.2 Identification	12
1.4.3 Estimation	15
1.5 Data	16
1.6 Empirical results	19
1.6.1 Effect of relaxing fiscal rules on policy outcomes	19
1.6.2 Validity tests	26
1.6.3 Political economy of fiscal adjustment	30
1.7 Conclusion	37
2 Ghost-House Busters: The Electoral Response to a Large Anti Tax Evasion Program	39
2.1 Introduction	39
2.2 The Ghost Buildings Program	44
2.3 Tax Enforcement and Retrospective Voting: A Conceptual Framework	47
2.4 Data and Descriptive Evidence	52
2.4.1 Data	52
2.4.2 The Correlates of Tax Evasion	55
2.4.3 The Political Determinants of the Ghost Buildings Registration	59
2.5 Empirical Strategy	61
2.5.1 The Electoral Response to the Ghost Buildings Program	61
2.5.2 Channels	67
2.6 Results	69

2.6.1	Baseline Results	69
2.6.2	Identification Validity and Robustness Checks	73
2.6.3	Channels	77
2.6.4	Alternative Explanations	82
2.7	Conclusion	84
3	Law, Economics and Culture: Theory of Mandated Benefits and Evidence from Maternal Leave Policies	87
3.1	Introduction	87
3.2	The Model	91
3.2.1	Setup	91
3.2.2	Gender Based Discrimination	92
3.2.3	Intolerance of Gender Based Discrimination	94
3.2.4	Optimal Length of Leave	96
3.2.5	The Effect of Attitudes on Leave	97
3.3	Empirical Analysis	97
3.3.1	Cross Country Correlation	97
3.3.2	Language and Attitudes	100
3.3.3	Maternity Leave and Language	105
3.3.4	Instrumental Variable Approach	107
3.4	Discussion	109
3.5	Conclusion	112
	Appendix A Appendix to Chapter 1	127
A.1	Supplementary Tables and Figures	127
	Appendix B Appendix to Chapter 2	139
B.1	Appendix Figures	140
B.2	Appendix Tables	142
	Appendix C Appendix to Chapter 3	145
C.1	Extension to Model	145
C.2	Data Appendix	146
C.3	Language Coding	147

List of Tables

1.1	The rules of the Domestic Stability Pact (DSP)	9
1.2	Legislative thresholds for Italian municipalities, 1997–2004	11
1.3	The rules of the Domestic Stability Pact (DSP)	16
1.4	Outcome variables, descriptive statistics	19
1.5	The effect of relaxing fiscal rules, diff-in-disc estimates	25
1.6	Falsification test in 1999	29
1.7	The political economy of deficit bias, part I	32
1.8	The political economy of deficit bias, part II	34
1.9	Fiscal restraints and budget management	36
2.1	Summary Statistics	56
2.2	The Determinants of Ghost Building Intensity (per 1,000 land parcels)	58
2.3	The Determinants of the Ghost Building Registration Rate	63
2.4	Ghost Building Intensity and Incumbent Reelection: Baseline Results	71
2.5	Ghost Building Intensity and Election Competitiveness	72
2.6	Ghost Building Intensity and Incumbent Reelection: Additional Controls	75
2.7	Ghost Building Intensity and Incumbent Reelection: Robustness	78
2.8	Ghost Building Registration and Incumbent Reelection	80
2.9	Ghost Building Intensity and Incumbent Reelection: Heterogeneity Analysis	82
2.10	Local Government Expenditures	83
3.1	Summary Statistics for Maternity Leave and Attitudes Regression	98
3.2	Maternity Leave and Attitudes	99
3.3	Personal Pronouns in English and Spanish	102
3.4	Summary Statistics for World Value Survey Regression	104
3.5	Language and Attitudes	105
3.6	Summary Statistics for Maternity Leave and Language Regression	106
3.7	Maternity Leave and Language	107
3.8	Instrumental Variable Approach	108
3.9	Difference between Maternity and Paternity Leave	111

A.1	Variables' description and sources	128
A.2	The effect of relaxing fiscal rules, estimates with covariates	130
A.3	Balance tests of time-invariant characteristics	131
A.4	Balance tests of potentially endogenous characteristics	132
A.5	The political economy of deficit bias, part I – Falsification test	133
A.6	The political economy of deficit bias, part II – Falsification test	134
A.7	Fiscal restraints and budget management – Falsification test	135
B.1	Variables' description and sources	143
B.2	Political Variables' description and sources	144
C.1	Number of Cases of Gender Differentiated Pronouns	147

List of Figures

1.1	Difference-in-discontinuities and Yearly RD coefficients	21
1.2	Difference-in-discontinuities for policy outcomes and tax instruments	23
1.3	Yearly RD estimates for policy outcomes and tax instruments	24
2.1	The Ghost Building Identification Process	46
2.2	Number of Elections per Year	54
2.3	<i>Ghost Building Intensity</i> (per 1,000 land registry parcels)	57
2.4	Registered Ghost Building Intensity	60
2.5	Ghost Building Registration Rate	62
2.6	Ghost Buildings Program Inception Year	66
2.7	Difference in reelection rates pre- to post- Ghost Buildings program	69
2.8	Difference in reelection rates pre- to post- <i>Placebo</i> Ghost Buildings program .	74
2.9	Ghost Building Intensity Coefficient by Election Pre/Post Program	76
3.1	The Labor Market for Men (left) and Women (right)	93
3.2	The Labor Market for Men (left) and Women (right) when Society is Intolerant of Discrimination	95
B.1	Ghost Building Intensity Coefficient by Election Pre/Post Program	140
B.2	Robustness to changes in election sample time span	141

Acknowledgments

I have been privileged to get to know, interact, and exchange ideas with excellent people during my years at Harvard. My advisors, colleagues, and friends have made this time challenging, productive, and enjoyable.

Alberto Alesina has impressed me since the very first moment I started to study economics. During my years at Harvard, Alberto always encouraged me to think about the big picture, and about important topics. He changed radically my ideas of the interactions between a student and a professor, and I will miss the discussions with him. Robert Barro gave me the resources and the confidence that allowed me to focus on my research in the last years of graduate schools, and he encouraged me in an uncountable number of ways. The discussions with him often turned a confused idea into a coherent research project, and proved to me how important is intellectual curiosity in the life of an academic. I was also lucky to have Raj Chetty as an advisor. The discussions with him and his insights have greatly improved my papers and challenged my way of thinking about economics. Raj has always been supportive even when there was no reason to be so. The approach in this thesis owes so much to what Larry Katz taught me in his labor economics class and during our interactions that I consider Larry my (unwilling) ghost-coauthor of some of the chapters. Without his inspiration, I probably would not have been able to arrive at this point.

I leave Harvard knowing that I will try to emulate my advisors in many ways, and this will surely be a goal impossible to attain.

I owe a big intellectual debt toward all my coauthors during those years. I learnt so much from them, and I benefited a lot from their unique perspective. Thanks to Fernanda Brollo, Lorenzo Casaburi, Yehonatan Givati, Veronica Grembi, Tommaso Nannicini, Giacomo Ponzetto, Andrea Stella and Guido Tabellini. I would also like to thank the professors that allowed me to continue my studies in the US by writing a letter supporting my application

to graduate schools and encouraging me to pursue the plan: thanks to Erio Castagnoli, Frank Diebold, Roberto Perotti and Guido Tabellini. Special thanks go also to Philippe Aghion and David Laibson. They are not in my thesis committee only because of the limit requirements.

At Harvard, I benefited from discussions with many people. The (non-exhaustive) list includes Davide Cantoni, David Cutler, Eduardo Dávila, Wenxin Du, Kathy Edin, Ed Glaeser, Nathan Hendren, Jisoo Hwang, Sandy Jencks, Tao Jin, Asim Khwaja, Michael Kremer, Pepe Montiel Olea, Joana Naritomi, Nathan Nunn, Mandy Pallais, Rohini Pande, Nico Perez-Truglia, James Robinson, Dana Rotz, László Sándor, Andrei Shleifer, Seth Stephens-Davidowitz, Theda Skocpol, Lucia Tian, Rodrigo Wagner, Bruce Western, David Yanagizawa-Drott.

Thanks, too and especially, to my friends and parents, for their questions and unquestioning support; and to Manisha, whose love has driven me to continually seek improvement as a researcher, and as a person.

Chapter 1

Policy Responses to Fiscal Restraints: A Difference-in-Discontinuities Design¹

1.1 Introduction

Can fiscal restraints imposed on local governments create incentives for reducing the accumulation of debt? The need for fiscal adjustment in the aftermath of the Great Recession has revived interest in fiscal rules aimed at disciplining the discretionary power of policy makers. Despite extensive research, the impact of fiscal restraints on debt accumulation and their effectiveness in reducing politically motivated deficit remain highly debated.² As the authors in the literature have acknowledged—see, for example, the discussion in Poterba (1996) or Alesina and Perotti (1996)—the search for a definitive conclusion is hampered by the potentially endogenous decision of whether to adopt fiscal rules or not.³

¹Co-authored with Veronica Grembi and Tommaso Nannicini.

²As indicated by Drazen (2002) in a review article: “A key question (perhaps the key question) about fiscal rules is whether they have the effect of slowing the growth of deficits.”

³Furthermore, there has been limited investigation on fiscal rules at the local government level, where forms of “hidden” public debt can grow and raise fears about the overall financial sustainability of a country. For

In this paper, we study the effect of relaxing fiscal restraints at the local government level. We first show quasi-experimentally that fiscal rules do matter for restraining the accumulation of debt and the fiscal adjustment is concentrated on revenues. We then give evidence suggesting that the adjustment is driven by cities with more political distortions. We overcome previous data and identification limitations by using a novel identification design.

Our setting is Italy, where the central government set a target on deficit reduction for all municipal governments in 1999—the so called “Domestic Stability Pact,” DSP henceforth—and relaxed it for municipalities below 5,000 inhabitants in 2001. This policy change allows us to combine two sources of variation, before/after 2001 and just below/above 5,000 inhabitants, and implement what we call a “difference-in-discontinuities” (or “diff-in-disc”) design. It is important to note that a (standard) cross-sectional Regression Discontinuity (RD) design would not allow us to identify the effect of relaxing fiscal rules in this setting. The fact that there is another policy, started in the 1960s and still in place, according to which mayors of cities above 5,000 residents receive a higher salary, implies that analyzing the discontinuity at 5,000 residents in any given year would not identify the effect of the DSP.

The intuition for our identification strategy is simple. The diff-in-disc estimator takes the difference between the cross-sectional discontinuity at 5,000 after 2001 (when both fiscal rules and the mayor’s salary show a jump) and the cross-sectional discontinuity at 5,000 before 2001 (when only the mayor’s salary shows a jump). We derive precise identifying assumptions under which this estimator can identify different types of average treatment effects in the neighborhood of the population threshold. The assumptions are more local than those required for a difference-in-differences strategy, because they need to hold just at the threshold of interest and not in the whole sample. We can therefore improve upon the previous literature in terms of internal validity by identifying the causal effect of relaxing

a review of the current state of the literature, see Glaeser (2012). On the one hand, fiscal rules imposed on local governments might be particularly effective because a higher level of government can credibly enforce punishment. On the other hand, local governments usually have limited autonomy in adjusting fiscal policy.

fiscal restraints on policy outcomes.⁴

The main rule established by the DSP imposed a gradual reduction of the “fiscal gap,” defined as the municipal deficit net of transfers and debt service. The rationale for the exemption of municipalities with less than 5,000 inhabitants in 2001 was to avoid burdening very small towns with onerous requirements, as they may be disadvantaged by economies of scale in managing the municipal government. The penalties put in place for not complying with the DSP included a cut in the annual transfers from the central government, a ban on new hires, and a cut on reimbursement and non-absenteeism bonuses. This means that, in the regulatory environment we study, there is a central authority that can collect standardized public accounts and enforce punishment for non-compliers.

A number of studies have argued that there are several reasons why fiscal rules might be ineffective in restraining fiscal policy (see, for example, the review by Alesina and Perotti, 1996, and Wyplosz, 2012).⁵ Additionally, the approval of the DSP was accompanied by widespread skepticism about its effectiveness, as Italy usually ranks last among OECD countries in ratings of law enforcement and government effectiveness (e.g., see Kaufmann, Kraay, and Mastruzzi, 2010). In light of these arguments, our results suggest that the lessons we draw from Italian municipalities on the effectiveness of fiscal rules may extend to other regulatory environments where the fiscal authority setting the rules faces critical ex-ante commitment problems.

We divide our empirical analysis into three parts. First, we analyze the impacts of relaxing fiscal restraints on the deficit, which is the main policy variable of interest, and on the fiscal gap, which is the main target of the law. We find that relaxing fiscal rules

⁴See Section 1.4 for a formal discussion of identification, estimation, and diagnostics in diff-in-disc.

⁵First, the fact that subnational policy makers have limited discretion in changing fiscal policy is the central reason for which fiscal rules on local governments might not work. Second, most fiscal rules, including those we are analyzing, are not embedded in the constitution. This implies that fiscal responsibility laws can be frequently changed and revised, and they might suffer from the same time inconsistency problem that characterizes fiscal policy. Additionally, Alesina and Perotti (1996) argue that lax enforcement is one of the reasons why fiscal rules might not work. This concern, however, is more relevant at the national level, since at the local level the central government can be a credible enforcement authority. Finally, rules usually target only some parts of the budget and this offers opportunities for policy makers to sidestep the rules by complying with them without changing the overall fiscal discipline—see Milesi-Ferretti (2003).

translates into a larger fiscal gap of about 40 to 60 percent over the course of the following 4 years. This large effect on the main target of the DSP has real consequences for policy outcomes, as unconstrained municipalities increase their deficit by 20 Euros per capita (2 percent of the total budget). The fact that we find an effect not only for the target of the DSP, but also for the main policy variable of interest not targeted by the law (that is, debt accumulation) alleviates concerns arising from the possibility of creative accounting.

In the second part of our empirical analysis, we study the composition of the fiscal adjustment, by analyzing municipal financial reports and administrative data on municipal tax rates, which are set by the local policy makers. We find that unconstrained municipalities have statistically similar expenditure levels with respect to constrained municipalities, but have lower tax revenues. This difference can be partially explained by the additional finding that municipalities for which fiscal rules are relaxed set lower tax rates. The main tax rates decided by Italian cities are a real estate tax rate on home property (*Imposta Comunale sugli Immobili*, ICI), which provides almost 50 percent of municipal tax revenues, and a surcharge on the personal income tax (*Imposta sul Reddito delle Persone Fisiche*, IRPEF), which amounts to about 10 percent of municipal tax revenues. Cities for which fiscal rules are relaxed have both a lower real estate tax rate (by about 14 percent) and a lower income tax surcharge (by about 30 percent) after the policy shift.

Finally, in the third part of the analysis, we exploit the fact that our setup—that is, an exogenously imposed fiscal destabilization—can provide new evidence on when fiscal restraints matter the most. On the one hand, the optimal tax smoothing theory would suggest that countercyclical deficits can increase welfare by equalizing the distortionary cost of taxation across booms and recessions.⁶ On the other hand, a persistent deficit bias might be the suboptimal result of the interplay between rational politicians, voters, and interest groups. Our empirical findings suggest that political factors play a first-order role in fiscal adjustments.

We first compare municipalities where only two political parties are represented in

⁶See Barro (1974), Barro (1979), and Lucas and Stokey (1983).

the local legislative assembly (about half of the sample) versus municipalities with more parties. Our results show that relaxing fiscal rules increases the deficit only if more than two parties are seated in the assembly, which must approve the budget proposed by the mayor. This finding is consistent with models that explain deficit as the result of political fragmentation and of dynamic common pool (see Persson and Tabellini, 2000), and also with the cross-country evidence that coalition governments are associated with higher deficits (see Roubini and Sachs, 1989; Kontopoulos and Perotti, 1999).

We then study if the relaxation of fiscal rules is affected by whether the mayor faces a binding term limit or not. We find that the increase in deficit bias arises only for mayors who can be reelected. This result is consistent with models linking deficit to reelection incentives (see Aghion and Bolton, 1990) or to politicians' pandering to voters (see Maskin and Tirole, 2004). We also show that cities that increase the municipal deficit after the relaxation of fiscal rules have an older population. These results are consistent with the model of Song, Storesletten, and Zilibotti (2012), according to which young citizens have a disciplining role for fiscal policy because they internalize the future costs of present fiscal instability.

Furthermore, we relate our findings to models that formalize the welfare costs and benefits of fiscal restraints. First, we show that the increase in deficit bias arises only for mayors that systematically underprovide the public good promised to the voters in the provisional budget. This evidence can be interpreted as consistent with the model of Besley (2007), to the extent to which politicians who consistently overpromise public goods are those who care more about reelection rather than social welfare. Finally, we relate our findings to the empirical prediction of Battaglini and Coate (2008). These authors predict that the welfare costs of restraining fiscal policy are limited when deficit is a historically persistent phenomenon, rather than a cyclical one. We show that our results are not driven by cities with cyclical deficit, and—if anything—the opposite is true, even if data limitations do not allow us to draw a definitive conclusion.

Our results survive a large number of robustness checks. Among those, we use the introduction of the DSP for all municipalities in 1999 as a falsification test to show that

our results are not driven by cities just below and above 5,000 responding differently to the same set of fiscal rules. This test suggests that better paid (and, hence, better selected) mayors do not react differently to the relaxation of the DSP compared to mayors with a lower salary, and it is reassuring for the external validity of our results. Additionally, we repeat our falsification exercise interacting our treatment variable with the heterogeneity dimensions discussed above, in order to show that the law did not bind differently across those subsamples.

The paper proceeds as follows. Section 1.2 summarizes the relevant literature. Section 1.3 describes the Italian institutional framework. Section 1.4 lays out our identification and estimation strategy. Section 1.5 describes the data. Section 1.6 discusses the empirical results and validity tests. We conclude with Section 1.7.

1.2 Related literature

This paper relates to two strands of literature. First, we contribute to the literature that has analyzed the effectiveness of fiscal rules.⁷ Indeed, a number of empirical studies have tried to evaluate whether fiscal rules are associated with sounder budget outcomes, reaching mixed conclusions. The evidence primarily comes from cross-country comparisons in specific regions, such as the European Union (see Hallerberg and Von Hagen, 1999; Debrun et al., 2008) or Latin America (see Alesina, Hausmann, Hommes, and Stein, 1996), and from local governments in a federal state, such as the U.S. (see Poterba, 1994; 1996).⁸ While some studies find that fiscal rules do indeed result in lower budget imbalances (e.g., see Knight, 2000), others stress the reasons for which fiscal restraints might not be effective (e.g., see

⁷For surveys, see Poterba and Von Hagen (1999), Rodden, Eskeland, and Litvack (2003), and Wyplosz (2012). For an extensive review on the types of rules and the main empirical evaluations of their impact, see IMF (2009). Balassone, Franco, and Zotteri (2004) review the literature on subnational fiscal rules in the European Union. As we focus on local governments, our results are particularly relevant for the literature that has emphasized that the implementation of subnational fiscal rules faces serious commitment problems, in the form of future overhaul, soft budget constraints, and lack of enforcement: see, among others, Eichengreen and von Hagen (1996), Braun and Tommasi (2004), Sutherland et al. (2005), and Ter-Minassian (2007).

⁸Studies on the U.S. include Von Hagen (1991), Alt and Lowry (1994), Bayoumi and Eichengreen (1995), Bohn and Inman (1996), Alesina and Bayoumi (1996), Auerbach (2006), and Fatas and Mihov (2006).

Alesina and Perotti, 1999).⁹ We provide a quasi-experimental design where the effectiveness of fiscal rules is evaluated by controlling for omitted factors that may affect previous results, such as the fact that more disciplined constituencies introduce tighter rules, or that (current and past) legal institutions are endogenous to cultural values.¹⁰

Secondly, we contribute to the large literature on the political economy of deficit determination, as we identify a set of politicians' and voters' characteristics associated with larger deficit bias.¹¹ From a normative perspective, it is not obvious whether tight rules such as a balanced budget requirement are optimal or not. As discussed above, restricting fiscal policy is not the first best in standard macroeconomic models without political distortions. However, restricting fiscal policy might become optimal when deficit is the suboptimal result of the interplay between rational politicians, voters, and interest groups (such as the aforementioned Alesina and Tabellini, 1990; Persson and Svensson, 1989; Aghion and Bolton, 1990; Besley, 2007; Battaglini and Coate, 2008; Song, Soresletten and Zilibotti, 2012).¹² These models deliver different predictions on the types of polities where one should expect the emergence of larger deficits. Our empirical findings shed new light on these dimensions.

1.3 Institutional framework

The Italian municipal government (*Comune*) is composed of a mayor (*Sindaco*), an executive committee (*Giunta*) that is appointed by the mayor, and an elected city council (*Consiglio Comunale*) that must endorse the annual budget proposed by the mayor. The mayor and the executive committee—whose members can be dismissed by the mayor at will—propose

⁹Knight (2000) uses the difficulty of amending U.S. state constitutions as an instrument for analyzing the effect of supermajority requirements for tax increases on tax rates.

¹⁰On the endogenous determination of laws, see Aghion, Alesina, and Trebbi (2004) and Givati and Troiano (2011). From a theoretical perspective, other authors analyze the welfare effect of fiscal restraints: Besley and Smart (2007) study limits on the size of government in a two-period agency model; Bassetto and Sargent (2006) study the welfare case for allowing the government to issue debt only to finance certain expenditures.

¹¹This literature has been reviewed by Alesina and Perotti (1999).

¹²Other political economy models on deficit determination include Tabellini and Alesina (1990), Lizzeri (1999), Azzimonti, Battaglini, and Coate (2008), and Yared (2010).

changes in fiscal policy, such as adjustments in the tax rates. Subsequently, the city council votes on the proposed changes. Since 1993, mayors are directly elected (with single round plurality rule in cities below 15,000 inhabitants) and face a two-term limit. Municipalities manage about 10 percent of total public expenditure and are in charge of a wide range of services, including water supply, waste management, municipal police, infrastructures, welfare, and housing. Only about 20 percent of revenues are local revenues.

After the European Union adopted its Stability and Growth Pact in 1997, some European countries—including Italy—adopted subnational fiscal rules to keep local governments accountable. In December 1998, the Italian annual budget law (*Legge Finanziaria*) for 1999 introduced a set of rules that constrained all municipalities in terms of fiscal discipline, the aforementioned Domestic Stability Pact or DSP (*Patto di Stabilità Interno*).¹³

Municipal governments were constrained to keep the growth of their fiscal gap—defined as deficit, net of transfers and debt service—under tight control. The rationale for the exclusion of debt service and transfers in the definition of the DSP target is twofold. First, mayors are not accountable for expenses on interests (which depend on previously contracted loans) and for revenues from transfers (which are not raised by the municipality). Second, these two items are the tools that the central government uses to enforce fiscal rules, reducing interest payments for compliers and cutting transfers for non-compliers. The punishment established for not complying with the DSP included the following penalties: (i) 5 percent cut in the annual transfers from the central government; (ii) ban on municipal hires; (iii) 30 percent cut on reimbursement and non-absenteeism bonuses for the employees of the municipal administration. Cities complying with the DSP, instead, benefited from a reduction of the expenses on interests for loans from the central government.¹⁴

¹³See Law 23 December 1998, no. 448, article 28.

¹⁴Consistently with the law, we compute fiscal gap with the formula: Fiscal Gap = (Total Expenditures - Debt Service) - (Total Revenues - Transfers). See the Appendix Table A1 for more details on the definition of policy outcomes. Unfortunately, the Ministry of the Interior does not release the list of municipalities that did not comply with the rule according to its records. As discussed in Section 1.6.1, we find suggestive evidence that the DSP penalties were enforced, as there is a correlation between non-compliance (as estimated in our data) and lower transfers (which are the main DSP enforcement mechanism).

The exact DSP rule constraining the fiscal gap changed from one year to another, but over our sample period it consisted in imposing a cap on the growth rate of the gap. Table 1.3 summarizes the evolution of the DSP over our sample period.

Table 1.1: *The rules of the Domestic Stability Pact (DSP)*

Year	Target of the DSP rules	Covered municipalities
1997	None	All
1998	None	All
1999	Fiscal gap: zero growth	All
2000	Fiscal gap: zero growth	All
2001	Fiscal gap: max 3% growth	Above 5,000
2002	Fiscal gap: max 2.5% growth	Above 5,000
2003	Fiscal gap: zero growth	Above 5,000
2004	Fiscal gap: zero growth	Above 5,000

Notes. The *Domestic Stability Pact* is a set of fiscal rules imposed by the central government to discipline the fiscal management of local governments. The main target is the *Fiscal gap* (see the Appendix Table A4 for details). The growth of fiscal gap with respect to its value two budget years before is constrained to be either zero or below 2.5%/3% depending on the year of the DSP. Legislative sources: annual national budget law (*Legge Finanziaria*) from 1999 to 2004.

The cap varies between a minimum of zero (no growth allowed) and a maximum of 3 percent, the benchmark being the fiscal gap two years before the actual budget year (this means that, for instance, the growth rate in 2004 is calculated with respect to the fiscal gap in 2002).

In evaluating the impact of the DSP on fiscal discipline, we therefore focus on the pattern of both deficit and fiscal gap. Constrained and unconstrained municipalities can accumulate debt, but if they run into fiscal distress they need to go through a special procedure of budget consolidation (*Piano di Risanamento*). One possible concern can be that relaxing fiscal rules induces expectations in our treated cities that they will be bailed out in case of situations of fiscal distress.¹⁵ While we acknowledge the possibility that changes in fiscal restraints can always be confounded with changes in expectations, both legal and anecdotal evidence are consistent with the view that the Italian government made substantive effort to keep expectations of bailing out as low as possible in the period of interest. In 2001, the

¹⁵In general, Italian municipalities can finance their debt through the emission of bonds (*Buoni Obbligazionari Comunali*) or with loans from a central administrative agency (*Cassa Depositi e Prestiti*) and from private banks.

Italian Constitution undertook a substantial revision that tried to introduce a higher degree of fiscal decentralization while making bailouts unconstitutional.¹⁶

After 2001, all municipalities below 5,000 inhabitants were exempted by the DSP.¹⁷ The motivation for this exemption was not made explicit by the central government, but it is probably linked to the goal of providing some relief to small municipalities in the presence of economies of scale in managing the municipal government. Fiscal rules, however, are not the only policy varying with population size at 5,000. In particular, at this cutoff, there is a sharp increase in the wage received by the mayor and by the other members of the executive committee, based on a remuneration policy that has been in place since the early 1960s. Gagliarducci and Nannicini (2012) show that the wage increase at 5,000 attracts more educated individuals into politics and improves their performance once elected. Table 1.2 summarizes all the Italian policies on municipal governments relying on population thresholds over our sample period.

Population size determines the size of the city council; the size of the executive committee; the electoral rule; and whether a municipality can have additional elective bodies at the neighborhood level. But only the DSP (after 2001) and the salary of local politicians display a discontinuity at the 5,000 cutoff.

In 2002, regions with special autonomy (*Regioni a Statuto Speciale*) were allowed to set their own fiscal rules for municipal governments, and this is why we do not consider these regions in our study. Furthermore, since 2005 fiscal rules have been frequently changing from one year to another, shifting the population cutoff from 5,000 to 3,000 and back, and replacing the fiscal gap requirement with expenditure caps in some years. This is the reason why we focus our empirical evaluation on the period from 1997 to 2004.

¹⁶The new article 119 of the Italian Constitution specifically forbids the increase of governmental transfers to local governments in fiscal distress. Anecdotal evidence confirms a hard-line stance by the central government toward indebted municipalities. For instance, Taranto, a medium sized Italian city, declared bankruptcy in 2006; local newspapers reporting on the fiscal situation of the city (e.g., see *Taranto Sera*) stressed how the city had to undertake a multi-year repayment plan, without any help from the central government, and, after six years, almost half of the debt was still outstanding; public services and wage of public employees were suspended for some months after the bankruptcy, and local tax rates were significantly raised.

¹⁷See Law 23 December 2000, no. 388, article 53.

Table 1.2: *Legislative thresholds for Italian municipalities, 1997–2004*

Population	Wage of mayor	Wage of executive committee	Size of executive committee	Size of city council	Electoral rule
Below 1,000	1,291	15%	4	12	single
1,000-3,000	1,446	20%	4	12	single
3,000-5,000	2,169	20%	4	16	single
5,000-10,000	2,789	50%	4	16	single
10,000-15,000	3,099	55%	6	20	single
15,000-30,000	3,099	55%	6	20	runoff
30,000-50,000	3,460	55%	6	30	runoff
50,000-100,000	4,132	75%	6	30	runoff
100,000-250,000	5,010	75%	10	40	runoff
250,000-500,000	5,784	75%	12	46	runoff
Above 500,000	7,798	75%	14-16	50-60	runoff

Notes. Policies varying at different legislative thresholds in the period 1999–2004. *Population* is the number of resident inhabitants as measured by the last available Census. *Wage of mayor* and *Wage of executive committee* refer to the monthly gross wage of the mayor and the members of the executive committee, respectively; the latter is expressed as a percentage of the former, which refers to 2000 and is measured in Euros. *Size of executive committee* is the maximum allowed number of executives appointed by the mayor. *Size of city council* is the number of seats in the city council. The wage thresholds at 1,000 and 10,000 were introduced in 2000; all of the other thresholds date back to 1960. Since 1993, the *Electoral rule* for the mayor is plurality with either single round or runoff.

In the next section, we explain how we exploit these unique Italian institutions to identify the effect of fiscal restraints on fiscal discipline.

1.4 Difference-in-discontinuities design

1.4.1 Setup

Define $Y_{it}(1)$ and $Y_{it}(0)$ as the potential policy outcomes of municipality i at time t in the case of treatment ($D_{it} = 1$) and no treatment ($D_{it} = 0$), respectively. Because of the institutions described above, the treatment D_{it} coincides with “relaxing fiscal rules.” If $t \geq t_0$, only municipalities below the population cutoff p_c are treated; the running variable p_i is set at the Census level and therefore time-invariant. Formally, treatment assignment is given by:

$$D_{it} = \begin{cases} 1 & \text{if } p_i \leq p_c, t \geq t_0 \\ 0 & \text{otherwise.} \end{cases} \quad (1.1)$$

Borrowing the notation from Hahn, Todd, and Van der Klaauw (2001), define $Z^- \equiv$

$\lim_{p \rightarrow p_c^-} E[Z_{it}|p_i = p, t \geq t_0]$ and $Z^+ \equiv \lim_{p \rightarrow p_c^+} E[Z_{it}|p_i = p, t \geq t_0]$, with $Z = Y(1), Y(0), D, Y$, where Y is the observed policy outcome. Hahn, Todd, and Van der Klaauw (2001) derive precise conditions under which the cross-sectional RD estimator after t_0 , defined as $\hat{\tau}_{RD} = Y^- - Y^+$, identifies the average treatment effect at the cutoff, $E[Y_{it}(1) - Y_{it}(0)|p = p_c]$. The identifying assumptions require D_{it} to be independent of $Y_{it}(1) - Y_{it}(0)$ conditional on p_i near p_c , and potential outcomes to be continuous at p_c : $Y(1)^- = Y(1)^+$ and $Y(0)^- = Y(0)^+$.

In our setting, as it is often the case when different policies share the same cutoff, the above continuity assumptions break down, because also politicians' wages sharply change at p_c . Define: $\gamma_1 \equiv Y(1)^- - Y(1)^+ \neq 0$; $\gamma_0 \equiv Y(0)^- - Y(0)^+ \neq 0$. Here, γ_1 and γ_0 capture the effects of the confounding policy discontinuity on potential outcomes. As a result:

$$\hat{\tau}_{RD} \equiv Y^- - Y^+ = Y(1)^- - Y(0)^+ = Y(1)^- - Y(0)^- + Y(0)^- - Y(0)^+ \equiv NATT + \gamma_0, \quad (1.2)$$

or equivalently

$$\hat{\tau}_{RD} \equiv Y^- - Y^+ = Y(1)^- - Y(0)^+ = Y(1)^+ - Y(0)^+ + Y(1)^- - Y(1)^+ \equiv NATU + \gamma_1, \quad (1.3)$$

where we define NATT as the Neighborhood Average Treatment effect on the Treated (i.e., on units below p_c) and NATU as the Neighborhood Average Treatment effect on the Untreated (i.e., on units above p_c).¹⁸ Only when $\gamma_1 = \gamma_0$, the two estimands are equal and represent the Neighborhood Average Treatment Effect (NATE), $E[Y_{it}(1) - Y_{it}(0)|p = p_c]$, which is the standard estimand identified in regression discontinuity design.

1.4.2 Identification

We now show how to overcome the identification problem discussed above. Information on the pre-treatment period ($t < t_0$) allows us to remove the bias under local assumptions. Analogously to the post-treatment period, define: $\tilde{Z}^- \equiv \lim_{p \rightarrow p_c^-} E[Z_{it}|p_i = p, t < t_0]$ and $\tilde{Z}^+ \equiv \lim_{p \rightarrow p_c^+} E[Z_{it}|p_i = p, t < t_0]$, with $Z = Y(1), Y(0), Y$. To identify the causal effect

¹⁸In words, the NATT (NATU) captures the effect of relaxing fiscal rules for municipalities just below (above) p_c . Because the running variable p_i is time-invariant, with a slight abuse of notation, we define observations below p_c as "treated" and above as "untreated" (although they are actually so only after t_0).

of relaxing fiscal rules, we exploit both the (sharp) discontinuous variation at p_c and the outcome time variation after t_0 :

$$\hat{\tau}_{DD} \equiv (Y^- - Y^+) - (\tilde{Y}^- - \tilde{Y}^+). \quad (1.4)$$

We call $\hat{\tau}_{DD}$ “difference-in-discontinuities” estimator (shortly, diff-in-disc), because it rests on the intuition of combining a difference-in-differences strategy and an RD design.

Alternative approaches in the literature have exploited the longitudinal nature of the data in a regression discontinuity setup, such as the fixed-effect RD estimator in Pettersson-Lidbom (2012), the first-difference RD estimator in Lemieux and Milligan (2008), or the dynamic RD design in Cellini, Ferreira, and Rothstein (2011). All of these estimators, however, are different from ours. In their setups, treatment assignment changes over time and identification rests on within-unit variation, while in our case the running variable is time-invariant. We are aware that other empirical studies implement some type of diff-in-disc strategy, but in this section we provide precise identification assumptions for the approach.¹⁹

Assumption 1 *The effect of the confounding policy discontinuity is constant over time: $Y(0)^- - Y(0)^+ = \tilde{Y}(0)^- - \tilde{Y}(0)^+$.*

Result 1 *Under Assumption 1, the diff-in-disc estimator identifies the NATT.*

Proof 1 $\hat{\tau}_{DD} \equiv (Y^- - Y^+) - (\tilde{Y}^- - \tilde{Y}^+) = (Y(1)^- - Y(0)^-) + (Y(0)^- - Y(0)^+) - (\tilde{Y}(0)^- - \tilde{Y}(0)^+) = Y(1)^- - Y(0)^- = NATT.$

Assumption 1 can be interpreted from two perspectives. First, it is similar to the RD assumption that potential outcomes are continuous, as it states that the *difference* in $Y_{it}(0)$ before and after t_0 is continuous at P_c . If the standard continuity assumption holds, Assumption 1 is also met and the diff-in-disc estimator should simply be used as a robustness

¹⁹Our econometric strategy also relates to evaluation designs that exploit the comparison between different discontinuities across space, such as in different U.S. states (see Dickert-Conlin and Elder, 2010) or for politicians facing different term limits (see Gagliarducci and Nannicini, 2013).

check for the standard RD design. But if the standard continuity assumption does not hold, as in our setting, Assumption 1 could still be met and provide the basis for identification under diff-in-disc. In Section 1.6, we validate the plausibility of this assumption by checking whether any manipulation of the running variable changes (or arises) over time.²⁰

From a second perspective, Assumption 1 requires the effect of the confounding policy discontinuity not to vary with time. In other words, it requires observations just below and just above p_c to be on a common trend. This is similar to the standard identifying assumption for difference-in-differences but is more local in nature, as it must be met only in a neighborhood of the policy cutoff. To test for (local) common trend, in Section 1.6, we estimate the pattern of discontinuities in Y_{it} before t_0 and we provide on how the regression discontinuity point estimates evolve over time.

Notice that the NATT is local in a twofold manner, because it involves the subpopulation hit by the interaction of two policies: the treatment and the confounding policy discontinuity. In our case, it captures the effect of relaxing fiscal rules for mayors with a lower wage. Under an additional homogeneity assumption, the diff-in-disc estimator identifies the NATE, which is the standard estimand in cross-sectional RD.

Assumption 2 *The effect of the confounding policy discontinuity is the same with and without treatment: $Y(1)^- - Y(1)^+ = Y(0)^- - Y(0)^+$.*

Result 2 *Under Assumptions 1 and 2, the diff-in-disc estimator identifies the NATE.*

Proof 2 *Because of Assumption 1, $\hat{\tau}_{DD} = \text{NATT}$. Because of Assumption 2, $\text{NATT} \equiv Y(1)^- - Y(0)^- = Y(1)^+ - Y(0)^+ \equiv \text{NATU}$. Therefore: $\text{NATT} = \text{NATU} = \text{NATE}$.*

Assumption 2 states that there must be no interaction between the treatment and the confounding policy discontinuity, and is similar to the additivity condition in difference-

²⁰Specifically, we extend the cross-sectional test of continuity of the density at p_c (see McCrary, 2008) to test for the continuity of the *difference* in the densities before and after t_0 . We also implement diff-in-disc estimations with time-invariant characteristics as outcomes, so as to indirectly test for changes in the pattern of manipulative sorting. As a further check in this direction, we include time-invariant characteristics and year fixed effects as covariates in the baseline diff-in-disc estimations; in the absence of manipulative sorting, point estimates are expected to remain similar and accuracy to increase.

in-differences (see Angrist and Krueger, 1998). In our institutional setting, this assumption would be violated if mayors just below and above p_c , who are paid differently, reacted to fiscal rules in a different way. In Section 1.6, under the maintained hypothesis that Assumption 1 holds, we directly test this second assumption exploiting the introduction of fiscal rules for all municipalities in 1999. In fact, if Assumption 2 holds, a falsification test implementing the diff-in-disc estimator in 1999 should deliver a zero effect.

1.4.3 Estimation

The diff-in-disc estimator can be implemented by estimating the boundary points of four regression functions of Y_{it} on p_i : two on both sides of p_c , before and after t_0 . We borrow two different estimation methods from the RD literature for this purpose: local linear regression and spline polynomial approximation.²¹

The first method fits linear regression functions to the observations distributed within a distance h on either side of p_c , both before and after t_0 . Formally, we restrict the sample to cities in the interval $p_i \in [p_c - h, p_c + h]$ and estimate the model:

$$Y_{it} = \delta_0 + \delta_1 P_i^* + J_i(\gamma_0 + \gamma_1 P_i^*) + T_t[\alpha_0 + \alpha_1 P_i^* + J_i(\beta_0 + \beta_1 P_i^*)] + \xi_{it}, \quad (1.5)$$

where J_i is a dummy for cities below 5,000, T_t an indicator for the post-treatment period, and $P_i^* = p_i - p_c$ the normalized population size. Standard errors are clustered at the city level. The coefficient β_0 is the diff-in-disc estimator and identifies the treatment effect of relaxing fiscal rules, as the treatment is $D_{it} = J_i \cdot T_t$. As suggested by Imbens and Lemieux (2008) and due to a lack of consensus on how to choose an optimal bandwidth, we present the robustness of our results to multiple bandwidths h .

The second method uses all observations and chooses a flexible functional form to fit the relationship between Y_{it} and p_i on either side of p_c , both before and after t_0 :

$$Y_{it} = \sum_{k=0}^q (\delta_k P_i^{*k}) + J_i \sum_{k=0}^q (\gamma_k P_i^{*k}) + T_t \left[\sum_{k=0}^q (\alpha_k P_i^{*k}) + J_i \sum_{k=0}^q (\beta_k P_i^{*k}) \right] + \xi_{it}. \quad (1.6)$$

²¹See Imbens and Lemieux (2008), Van der Klaauw (2008), and Lee and Lemieux (2010).

Again, standard errors are clustered at the city level, and the coefficient β_0 is the diff-in-disc estimator identifying the treatment effect of relaxing fiscal rules. We present the robustness of our results to multiple orders of the polynomial approximation (q).

1.5 Data

We use administrative data from the Italian Ministry of the Interior (*Ministero dell'Interno*) containing information at the municipality level on the universe of municipal financial reports, municipal tax rates, electoral outcomes, and individual characteristics of the mayor. Based on the local nature of our diff-in-disc design, we restrict the sample to Italian municipalities between 3,500 and 7,000 inhabitants.²²

Table 1.3: *The rules of the Domestic Stability Pact (DSP)*

Year	Target of the DSP rules	Covered municipalities
1997	None	All
1998	None	All
1999	Fiscal gap: zero growth	All
2000	Fiscal gap: zero growth	All
2001	Fiscal gap: max 3% growth	Above 5,000
2002	Fiscal gap: max 2.5% growth	Above 5,000
2003	Fiscal gap: zero growth	Above 5,000
2004	Fiscal gap: zero growth	Above 5,000

Notes. The *Domestic Stability Pact* is a set of fiscal rules imposed by the central government to discipline the fiscal management of local governments. The main target is the *Fiscal gap* (see the Appendix Table A4 for details). The growth of fiscal gap with respect to its value two budget years before is constrained to be either zero or below 2.5%/3% depending on the year of the DSP. Legislative sources: annual national budget law (*Legge Finanziaria*) from 1999 to 2004.

For the reason discussed in Section 1.3, we drop municipalities in regions with special autonomy. This leaves us with a final sample of 1,050 municipalities for a total of 6,300 observations from 1999 to 2004. Among them, 555 municipalities are treated after 2001 (because they are below 5,000 inhabitants) and 495 are in the control group. Our sample

²²We restrict the sample to the interval 3,500–7,000 to stay relatively far from the 3,000 threshold, where other policies change (see Table 1.2), and to balance the sample size on either side of the 5,000 threshold. All the results are robust to this interval choice, i.e., they are virtually unchanged for alternative choices, such as 3,250–6,750; 3,000–7,000; 3,500–6,500; 4,000–6,000; and 3,500–7,500 (available upon request).

contains about 13 percent of all Italian municipalities and about 8 percent of the national population.

The population size that decides treatment status is the 2001 Census. Because the relaxation of the DSP was decided in December 2000, it is very unlikely that municipalities had the time to influence their population and sort below the 5,000 threshold, and—on top of this—it is also unlikely that elected officials wanted to do that at the price of cutting their wage. In any case, in Section 1.6.2, we formally test for manipulative sorting below 5,000 before/after 2001 by comparing population size in the 1991 and 2001 Census.

The main variables of interest are the municipal financial report's categories. To measure fiscal discipline, we evaluate the deficit (total expenditures minus total revenues) and the fiscal gap (total expenditures minus total revenues, net of transfers and debt service), which is the target of the DSP. We divide expenditures into current outlays (including personnel expenditure), capital outlays (mostly investments), and debt service; and we divide revenues into municipal taxes, fees and tariffs, transfers from the central government, and other revenues. The main tax instruments decided by municipal governments are the real estate tax rate on home property (ICI), providing about 50 percent of their tax revenues, and the municipal surcharge on the personal income tax (IRPEF), amounting to about 10 percent of tax revenues.²³ See the Appendix Table A1 for precise definitions and data sources of all variables from municipal financial reports.

One possible concern in evaluating the reaction of policies and tax instruments to fiscal rules might be that mayors have very little autonomy in adjusting local revenues or expenditure, but this is not the case for Italian municipalities. On the revenues side, over our sample period, mayors could vary ICI within a bracket from 0.4 to 0.7 percent of the legal home value, and the IRPEF surcharge within a bracket from 0 to 0.5 percent of taxable income.²⁴ And they were also free to set other local taxes (such as those on building rights

²³Bordignon, Nannicini, and Tabellini (2012) also use ICI as the main policy tool of Italian municipalities.

²⁴One additional concern can be that mayors comply with the rule by simply manipulating legal home value. However, legal home value is not determined or updated by mayors, as indicated by the *DPR* 22 December 1986, no. 917. Only in 2005, not in our sample, the Law 23 December 2005, no. 266 gave to municipalities some

or the occupation of public areas), or fees and tariffs for the services they provided (such as waste management or child care). Additionally, Italian towns are characterized by a sizable level of tax evasion, which the mayor can decide to fight.²⁵ On the expenditures side, municipalities also have room for adjustment because about one third of the expenditures are classified as not rigid (that is, not attributable to payroll expenses and debt service). For instance, one way to reduce expenditures without affecting the level of services is outsourcing (e.g., child care provided by private firms with more labor flexibility and lower costs although the financing remains public). Furthermore, Bandiera, Prat, and Valletti (2009) show how similar Italian municipalities can pay very differently for similar goods, and they interpret this as evidence of passive waste. This implies that, even if all current expenditures were rigid (and this is not certainly the case), mayors would still have the ability to reduce passive waste in order to adjust the fiscal gap.

Our dataset also contains time-invariant information on each municipality (geographic location, area size in km², sea level in meters), as well as time-varying information on the elected mayor (age, years of schooling, tenure in office, term limit), on the socio-economic environment (taxable income of resident inhabitants, age structure of the population), and on the political environment (number of political parties seating in the city council). See again the Appendix Table A1 for their description and sources.

Table 1.4 provides descriptive statistics on the main outcome variables (policy outcomes and tax instruments) for cities below and above 5,000 inhabitants. All variables are per capita and expressed in real terms (with 2009 as base year); tax rates are in percentage points. Municipalities below (above) 5,000 manage an annual budget equal to almost 1,041 (943) Euros per capita in terms of expenditures, and the deficit amounts to about 15 (11) Euros. Taxes are only slightly lower than 20 percent of total revenues and higher in municipalities above 5,000. The main tax rates on ICI and the IRPEF surcharge, however, are fairly similar

weak power of requesting the update of the assessed value of the real estate tax base.

²⁵Casaburi and Troiano (2012) find that in 2007 over 2 millions of Italian buildings were not registered in the cadastral maps and thus were not part of the tax base for real estate and income tax.

for municipalities in the two groups.

Table 1.4: *Outcome variables, descriptive statistics*

	Municipalities above 5,000	Municipalities below 5,000
A. Fiscal discipline		
Deficit	11.080	15.457
Fiscal gap	170.724	208.624
B. Expenditures		
Current outlays	475.312	502.181
Capital outlays	438.838	508.794
Debt service	29.139	30.107
C. Revenues		
Taxes	194.887	175.825
Fees & tariffs	56.601	58.938
Central transfers	188.783	223.274
Other revenues	491.938	567.589
D. Tax instruments		
Real estate tax rate	0.587	0.576
Income tax surcharge	0.309	0.309
Obs.	2,970	3,330

Notes. Municipalities between 3,500 and 7,000 inhabitants; budget years between 1999 and 2004. The average values of per-capita policy outcomes are in 2009 Euros. The real estate tax rate and the income tax surcharge are in percentage points; the former can vary from 0.4 to 0.7 percent; the latter can vary from 0 to 0.5 percent.

As a benchmark, note that applying a difference-in-differences strategy to our dataset delivers the expected result: relaxing fiscal rules increases the deficit by 6.276 Euros and the fiscal gap by 48.278 Euros in a specification without municipality fixed effects, and by 5.279 and 16.669 with fixed effects, where both estimates on deficit are statistically significant at a 5 percent level and those on fiscal gap at a 1 percent level. However, in the next section, we discuss the results of our diff-in-disc design, which provides more credible evidence on the impact of relaxing fiscal rules on fiscal discipline for the reasons discussed above.

1.6 Empirical results

1.6.1 Effect of relaxing fiscal rules on policy outcomes

Table 1.5 contains the main (diff-in-disc) estimation results. For each outcome variable, we show the robustness of the results to four estimation methods: local linear regression

as in equation (1.5) with two different bandwidths (i.e., 500 and 750); spline polynomial approximation as in equation (1.6) with two different orders of the polynomial (i.e., 3rd and 4th).²⁶ The main outcomes of interest are the two measures of fiscal discipline: deficit and fiscal gap (see panel A of the table). While the latter is the main target of the DSP, we believe that the former should be the real variable of policy interest.

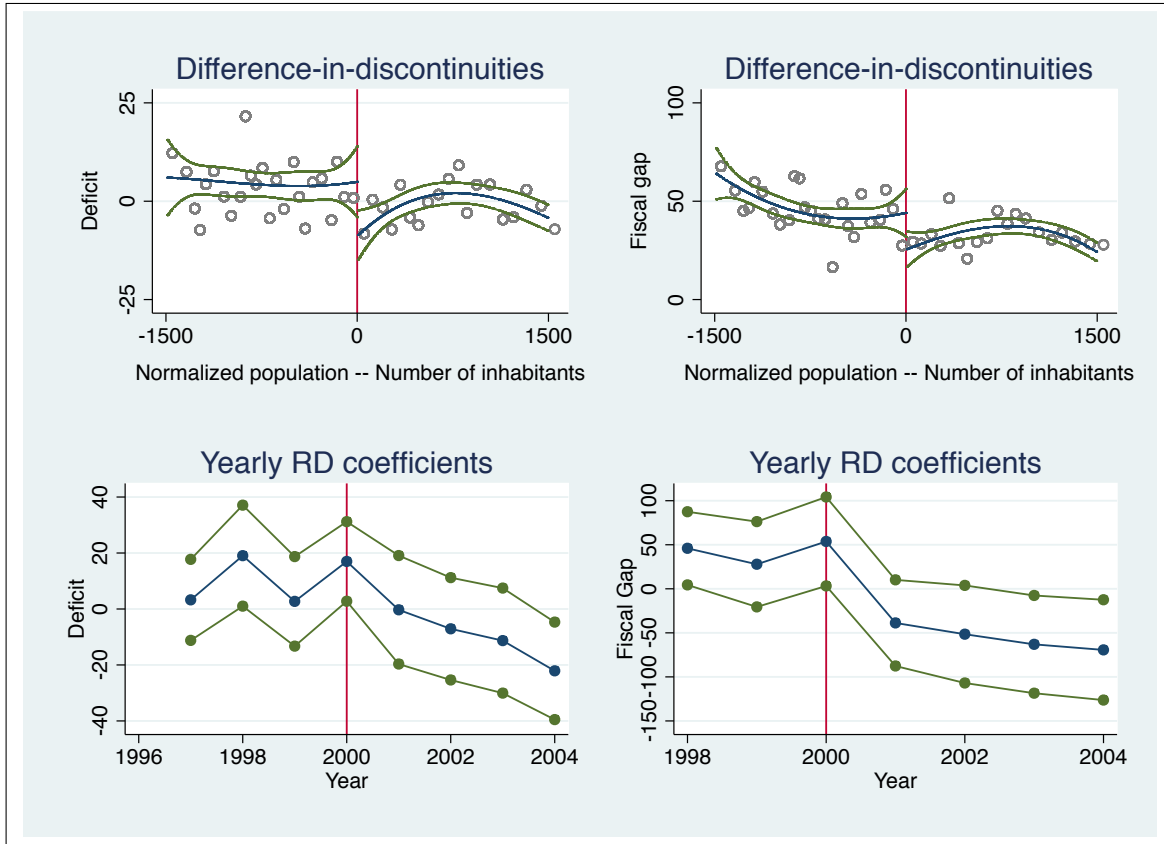
The impact of relaxing rules on the deficit is positive and significant both in statistical and in economic terms. The DSP relaxation increases deficit by about 20 Euros per capita with respect to a baseline situation of balanced budget. The deficit bias created by the relaxation of the DSP is also substantial from an economic point of view, as it amounts to about 2 percent of total expenditures. This effect is driven by a higher fiscal gap of about 40 to 60 percent, depending on the empirical specification. Both these effects are statistically significant at standard levels in all specifications, and the point estimates of the deficit are somehow more stable than those of fiscal gap.

These estimation results on fiscal discipline are consistent with the descriptive graphs shown in Figure 1.1 , where in the top panel we draw scatters and (3rd-order) polynomial fits of the differences between each post-2001 outcome value and each pre-2001 value. These graphs allow us to see whether those differences exhibit a discontinuity at the 5,000 threshold. We see that both variables measuring fiscal discipline exhibit a sharp jump when moving from the left to the right of the threshold in the whole sample (top left graph in both figures). Furthermore, in the other graphs in the bottom panel, we shed some light on the timing of the effect to provide evidence that high and low paid mayors were on parallel trends around the neighborhood of the 5,000 threshold. The evidence is consistent for both deficit and fiscal gap, as there is a change in the slope of the coefficients only after 2001. The observed discontinuities, however, remain statistically significant for all years.²⁷

In panels B and C of Table 1.5, we assess whether the fiscal (de)stabilization takes place

²⁶Results are robust to the use of additional bandwidths (i.e., 250 and 1,000) or additional orders of the polynomial (i.e., 2nd and 5th) and are available upon request. We use the two estimation methods that are standard in the RD literature. However, to address concerns about the sensitivity of our results to functional form assumptions, we also repeated the analysis implementing a simple t-test of the difference-in-discontinuities in closed intervals around the threshold (with intervals getting smaller and smaller) and we always obtained a

Figure 1.1: *Difference-in-discontinuities and Yearly RD coefficients*



Notes. For the Difference-in-discontinuities graphs. Vertical axis: difference of each post-rule (i.e. 2001, 2002, 2003, and 2004) outcome value and each pre-rule (i.e. 1999 and 2000) outcome value. Horizontal axis: actual population size minus 5,000. The central line is a spline 3rd-order polynomial fit; the lateral lines represent the 95% confidence interval. Scatter points are averaged over intervals of 50 inhabitants. For the Yearly RD coefficients. Vertical axis: point estimate of a LLR regression with optimal bandwidth computed following the Calonico, Cattaneo and Titiunik (2012) algorithm. Horizontal axis: year. The central line is the point estimate; the lateral lines represent the 95% confidence interval.

on the side of revenues or expenditure. While we cannot find a statistically significant impact on current outlays, capital outlays, and debt service, we find that tax revenues are lower by 20 to 45 percent in unconstrained municipalities (with respect to the average value of the control group and depending on the specification). Lower tax revenues are the result of lower tax rates decided by the municipal government (see panel D of Table 1.5). Cities for which fiscal rules are relaxed have a lower real estate tax rate by 14 percent and a lower income tax surcharge by 30 percent. Other revenues do not seem to be affected by the relaxation of fiscal restraints.²⁸

In Figure 1.3 we confirm that the common trend assumption is satisfied also for all our other financial reports' items.

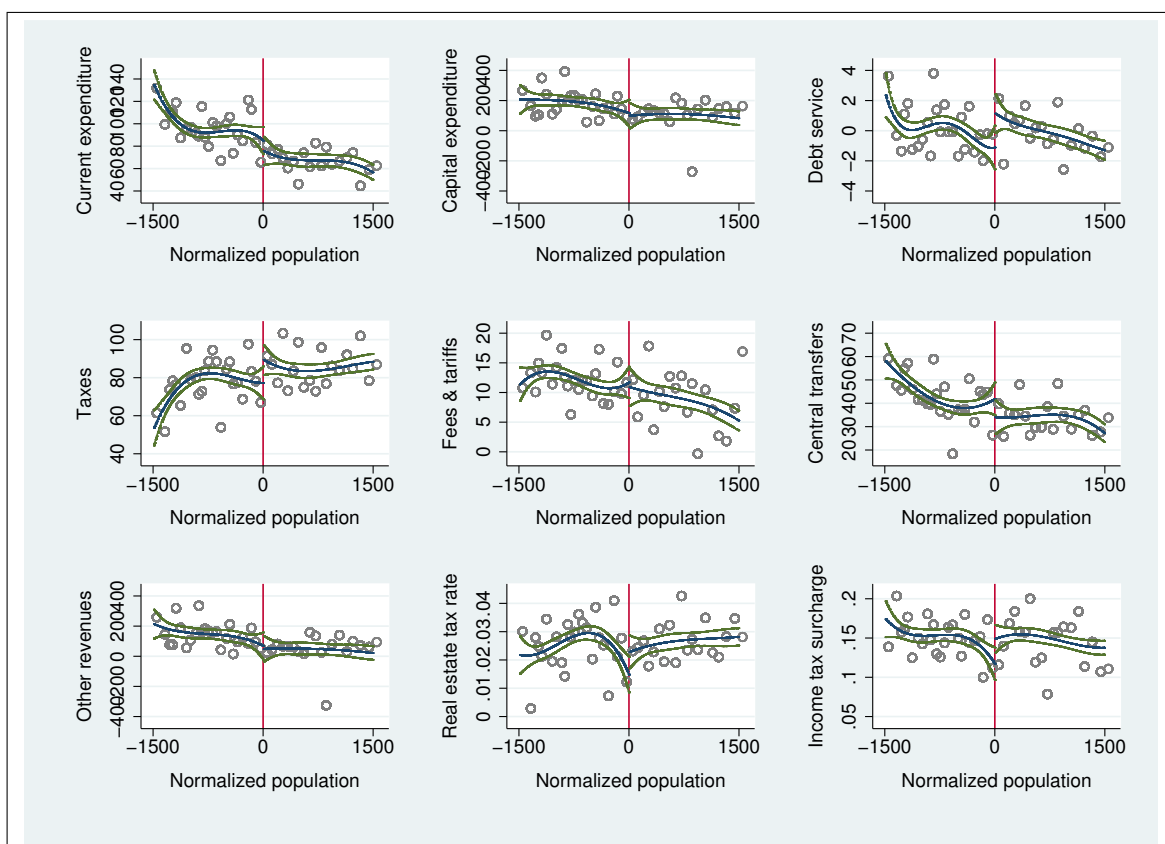
Also on the side of revenues, central transfers seem to be higher for unconstrained municipalities, although point estimates are not always statistically significant. This result cannot explain the above impact of relaxing fiscal rules on fiscal discipline because it goes in the opposite direction (that is, local governments running higher deficits receive larger transfers), and it is consistent with the design of the law, which allows the central government to cut transfers as an enforcement mechanism. This conjecture is consistent with our data. Although the Italian government did not release the official list of complying and non-complying municipalities, we can estimate compliance status in every year by applying the official rule summarized in Table 1.3 to our data. We find that complying municipalities amount to 68 percent of the total, and non-complying municipalities are also present around the 5,000 threshold, where the estimated compliance status shows a sharp discontinuity

difference statistically different from zero (results available upon request).

²⁷The yearly diff-in-disc estimates for 2001, 2002, 2003, and 2004 confirm the above graphical evidence on the timing of the effect of relaxing fiscal rules on fiscal discipline (available upon request).

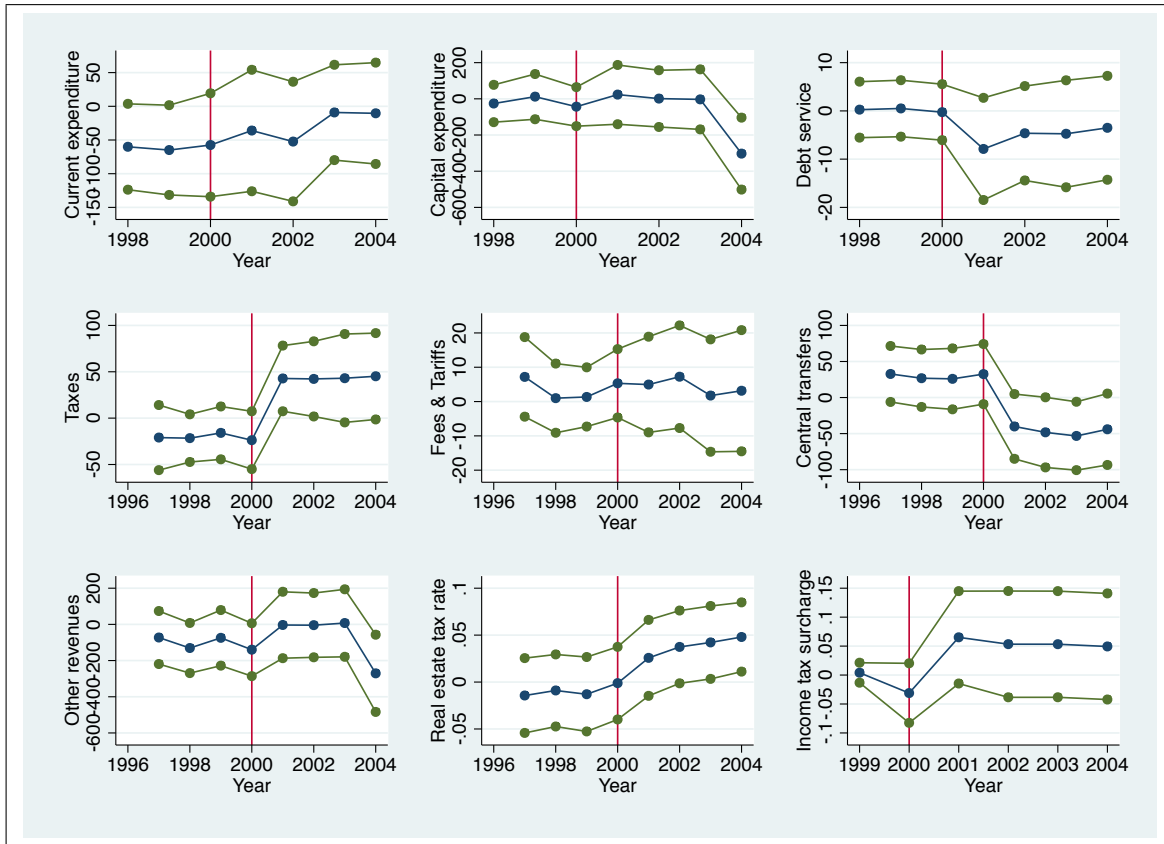
²⁸Other revenues include transfers from the European Union, other transfers, mortgages from administrative agencies, revenues coming from private properties owned by the municipality. Even if the standard errors for other revenues are bigger than the rest of our variables, visual inspection of the corresponding graph in Figure 1.2 reveals that standard errors are driven up by an outlier in this category. Repeating our analysis without this outlier consistently reduces the standard errors without affecting the other outcomes (results available upon request).

Figure 1.2: *Difference-in-discontinuities for policy outcomes and tax instruments*



Notes. Vertical axis: difference of each post-2001 (i.e. 2001, 2002, 2003, and 2004) outcome value and each pre-2001 (i.e. 1999 and 2000) outcome value. Horizontal axis: actual population size minus 5,000. The central line is a spline 3rd-order polynomial fit; the lateral lines represent the 95% confidence interval. Scatter points are averaged over intervals of 50 inhabitants.

Figure 1.3: Yearly RD estimates for policy outcomes and tax instruments



Notes. Vertical axis: point estimate of a LLR regression with optimal bandwidth computed following the Calonico, Cattaneo and Titiunik (2012) algorithm. Horizontal axis: year. The central line is the point estimate; the lateral lines represent the 95% confidence interval.

Table 1.5: *The effect of relaxing fiscal rules, diff-in-disc estimates*

	LLR $h = 500$	LLR $h = 750$	Spline poly 3 rd	Spline poly 4 th
A. Fiscal discipline				
Deficit	17.358** (8.377)	19.234*** (6.256)	21.449** (9.485)	25.109** (12.756)
Fiscal gap	61.748* (32.584)	57.839** (25.016)	102.202*** (38.463)	108.128** (48.372)
B. Expenditures				
Current outlays	-52.570 (55.367)	-4.208 (36.093)	-32.366 (55.631)	-60.520 (77.460)
Capital outlays	42.331 (87.304)	44.482 (63.528)	91.321 (103.145)	202.679 (140.024)
Debt service	-1.851 (6.895)	-1.593 (4.176)	-2.338 (6.631)	-2.375 (9.675)
C. Revenues				
Taxes	-45.248* (25.980)	-36.779* (19.185)	-57.028** (27.193)	-85.077** (35.162)
Fees & tariffs	-3.359 (10.214)	0.100 (7.416)	1.173 (10.601)	-4.051 (13.910)
Central transfers	42.539 (30.020)	37.012 (23.104)	78.414** (35.334)	80.644* (43.517)
Other revenues	-23.380 (108.275)	19.114 (72.838)	12.608 (118.470)	123.159 (165.666)
D. Tax instruments				
Real estate tax rate	-0.040* (0.024)	-0.028 (0.018)	-0.056** (0.026)	-0.060* (0.033)
Income tax surcharge	-0.036 (0.036)	-0.057** (0.029)	-0.058 (0.041)	-0.111** (0.051)
Obs.	2,080	3,068	6,300	6,300

Notes. Municipalities between 3,500 and 7,000 inhabitants; budget years between 1999 and 2004. Diff-in-disc estimates of the impact of relaxing fiscal rules on policy outcomes and tax instruments below 5,000 after 2001. Estimation methods: Local Linear Regression (LLR) with bandwidth $h = 500$ or $h = 750$, as in equation (1.5); spline polynomial approximation with 3rd-order or 4th-order polynomial, as in equation (1.6). All policy outcomes are per capita and in 2009 Euros. Tax instruments are in percentage points. Robust standard errors clustered at the municipality level are in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

of about 40 percent.²⁹ Future transfers appear to be strongly correlated with compliance status: in a specification that controls for year fixed effects, central transfers are larger by about 10.329 Euros per capita (standard error, 3.303) for complying municipalities.³⁰ This evidence is consistent with the institutional details discussed in Section 1.3, according to

²⁹Specifically, if we repeat our RD estimations using compliance status as a dependent variable, we obtain the following results for local linear regression ($h = 500$, $h = 750$) and spline polynomial approximation (3rd and 4th order), respectively: 0.450 (standard error, 0.070); 0.443 (0.054); 0.448 (0.076); 0.436 (0.096).

³⁰Results are robust to the use of municipality fixed effects.

which central transfers are used as the main enforcement device of the DSP.

Figure 1.2 provides graphical evidence on the diff-in-disc jumps at the 5,000 threshold in the policy outcomes and tax instruments. Consistently with the estimation results, tax revenues, ICI, and IRPEF show significant and negative jumps moving from just above to just below the 5,000 residents threshold. The graphs on expenditures confirm that these variables are reasonably stable across the threshold, even if the size of the standard errors in Table 1.5 does not allow us to reach a definitive conclusion on the possibility that a fraction of the adjustment also takes place through expenditures.

The previous results suggest that the fiscal adjustment mainly takes place on revenues. There are (at least) two possible explanations that can rationalize this finding. On the one hand, politicians might have a hard time convincing interest groups to cut expenditures, while taxpayers are more prone to internal free-riding and do not self-organize (see Olson, 1965). On the other hand, tax increases might be less salient than expenditure cuts for individuals (see Chetty, Looney, and Kroft, 2009) and thus they might be easier to adopt (or revert) for politicians.

1.6.2 Validity tests

As discussed in Section 1.4, the above estimation results rest on Assumption 1 and Assumption 2 for the identification of different average treatment effects in the neighborhood of the population threshold. In this section, we indirectly evaluate Assumption 1 by means of testing procedures aimed at detecting changes in manipulative sorting before/after 2001, and we directly test Assumption 2 in the falsification test that uses pre-treatment data.

The Appendix Figure A1 tests the null hypothesis of continuity of the *difference* in the density at 5,000 between the 1991 and the 2001 Census (top graph), by drawing both scatters and (3rd-order) polynomial fits. If mayors were able to manipulate population size and sort below the threshold to avoid fiscal rules, our estimates would still suffer from the selection bias that was common in the previous empirical literature. However, in principle, there is very little room for differential manipulation between the two Censuses, because (i) the

DSP is only enacted in December 2000, (ii) the Census is run independently by the National Statistical Office, so that false reporting should be ruled out, and (iii) mayors willing to sort below 5,000 to enjoy a relaxation of fiscal rules would pay the price of cutting their wage. Nevertheless, it might still be the case that some municipalities under financial stress tried to sort below 5,000 moving from the 1991 to the 2001 Census, by forcing some residents to leave or (more plausibly) not counter-reacting to population drops. Yet, the top graph in Figure A1 is reassuring about the absence of manipulation, as there is no jump in the difference between the two densities. The point estimate from the spline polynomial approximation is equal to -0.078 (standard error, 0.114) and therefore is not statistically different from zero. For the sake of completeness, we also report the cross-sectional density tests for 1991 (bottom left) and 2001 (bottom right). Also there, there is no evidence of manipulation.³¹

Furthermore, in the Appendix Table A2, we check for the balancing of time-invariant characteristics by including covariates, together with year fixed effects, in the baseline diff-in-disc estimations; as expected, point estimates remain almost unchanged and accuracy increases. The Appendix Table A3 further evaluates the absence of manipulation. We implement diff-in-disc estimations with time-invariant characteristics (geographic location, area size, and sea level) as outcome variables, but we use changing population numbers: the 1991 Census before the treatment year, and the 2001 Census afterwards. This is meant to assess whether the fraction of cities with certain fixed characteristics just below or above 5,000 varies from 1991 to 2001. No time-invariant characteristics display a statistically significant jump.³² We think that geographical location is a particularly interesting dimension here,

³¹The 1991 point estimate is 0.068 (0.082); the 2001 point estimate is -0.010 (0.076).

³²As an additional check, in the Appendix Table A4, we report balance tests of potentially endogenous characteristics. We implement diff-in-disc estimations with some (time-varying) economic or political characteristics of the municipality as outcome variables, using the 2001 Census population as the running variable as in the baseline specifications for the main policy outcomes. The time-varying characteristics we control for are the taxable income at the municipality level; the mayor's gender, years of schooling, and previous tenure in office; as well as the dimension of heterogeneous treatment effects we study in the next section (namely, term limit, number of parties, young cohorts, and speed of public good provision). These outcomes are potentially endogenous to the DSP, but detecting significant effects would disclose unexpected channels of adjustment through income, political selection, or public good delivery. This does not seem to be the case, as also these potentially endogenous variables are balanced around 5,000 before/after 2001.

because Italian geography is correlated with economic development, crime rates, labor market shirking, or political accountability (e.g., see Ichino and Maggi, 2000; Nannicini, Stella, Tabellini, and Troiano, 2012), and it could thus be associated with opportunistic manipulation too.

Based on this large amount of supporting evidence on Assumption 1, in Table 1.6 we directly test for Assumption 2 under the maintained hypothesis that Assumption 1 holds. In particular, we check whether cities just below or just above the 5,000 threshold respond differently to fiscal rules. We use the introduction of the DSP in 1999 for *all* municipalities as an experiment to test for the absence of any differential response around 5,000. Specifically, we implement diff-in-disc estimations in the interval 1997–2000, using 1999–2000 as the post-treatment period and 1997–1998 as the pre-treatment period.³³ All outcome variables are perfectly balanced around the threshold before/after 1999, confirming the assumption that there is no interaction between the DSP and the confounding wage discontinuity.³⁴

Finally, we perform a set of placebo tests to evaluate the possibility that our results arise from random chance rather than a causal relationship. In the Appendix Figure A2 and Figure A3, in the spirit of DellaVigna and La Ferrara (2012), we implement—respectively for deficit and for fiscal gap—a set of diff-in-disc estimations at false population thresholds below and above the 5,000 threshold (namely, any point from 4,900 to 4,400 and from 5,100 to 5,600 in order to stay sufficiently away from the true policy threshold). At these false thresholds, we expect to find no systematic evidence of treatment effects similar to our baseline results. The two figures report the cumulative density function of these 1,000 placebo point estimates (using a specification with 3rd-order spline polynomial), normalized

³³The city of Romentino was an outlier due to a lucrative sale of land in 1998 and it was removed from the sample. Our results do not change with the inclusion of this city, with the exception of bigger standard errors for other revenues (available upon request).

³⁴This falsification test suggests that the rule did not bind differently across the different sides of the population cutoff. We perform an additional robustness check by repeating yearly diff-in-disc estimations considering the last pre-treatment year, 2000, as the baseline year. As expected, none of the pre-treatment coefficients, from 1997 to 1999, is statistically different from the 2000 coefficient, while there is a sizable jump of the post-treatment coefficients (i.e., 2001, 2002, 2003, and 2004). This robustness check indirectly controls also for anticipation effects and for idiosyncratic variation in each single year (available upon request).

Table 1.6: *Falsification test in 1999*

	LLR $h = 500$	LLR $h = 750$	Spline poly 3 rd	Spline poly 4 th
A. Fiscal discipline				
Deficit	1.303 (7.604)	-4.735 (7.402)	0.942 (8.997)	-1.520 (10.673)
Fiscal gap	2.425 (11.158)	5.151 (8.836)	12.440 (11.882)	-2.693 (15.294)
B. Expenditures				
Current outlays	-2.027 (9.365)	-8.305 (8.724)	-11.384 (10.605)	-6.307 (11.584)
Capital outlays	0.528 (52.206)	-35.770 (45.759)	-13.273 (59.126)	-93.598 (98.401)
Debt service	-0.220 (1.239)	-0.347 (1.058)	-0.786 (1.278)	0.368 (1.591)
C. Revenues				
Taxes	-1.300 (4.024)	1.905 (3.668)	-3.411 (4.270)	-2.401 (4.923)
Fees & tariffs	3.168 (3.267)	2.681 (3.222)	-0.654 (3.428)	0.241 (3.490)
Central transfers	-2.270 (6.449)	1.755 (5.415)	-3.298 (7.187)	-11.149 (9.013)
Other revenues	34.986 (38.066)	-11.711 (32.788)	27.041 (43.639)	-62.725 (86.491)
D. Tax instruments				
Real estate tax rate	0.005 (0.009)	0.009 (0.007)	0.002 (0.009)	0.002 (0.011)
Obs.	1,260	1,848	4,176	4,176

Notes. Municipalities between 3,500 and 7,000 inhabitants; budget years between 1997 and 2001. Diff-in-disc estimates of the (false) impact of introducing fiscal rules on policy outcomes below 5,000 after 1999 (when no discontinuity was introduced by the DSP; see Table 1.3). Estimation methods: Local Linear Regression (LLR) with bandwidth $h = 500$ or $h = 750$, as in equation (1.5); spline polynomial approximation with 3rd-order or 4th-order polynomial, as in equation (1.6). All policy outcomes are per capita and in 2009 Euros. The real estate tax rate is in percentage points (the income tax surcharge is not available for this test because it was introduced in 1999). *Fiscal gap*, *Current outlays*, *Capital outlays*, and *Debt service* are not available in 1997; for these variables the observations in the four estimations, respectively, are: 945; 1,389; 3,135; 3,135. Robust standard errors clustered at the municipality level are in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

with respect to our baseline point estimates for deficit and fiscal gap. This means, for instance, that a normalized coefficient of 80 stands for a placebo point estimate equal to 80 percent of the true baseline estimate at 5,000. The intuition here is that we should not observe too many normalized coefficients outside the interval from -100 to +100. Indeed, all of the placebo coefficients are below our estimated coefficients for both deficit and fiscal gap, and the cumulative density function of the normalized coefficients is much steeper around zero. Only 0.8 percent of the normalized placebo coefficients for deficit is bigger

than the true coefficient in absolute value, while none of them is so for fiscal gap. On the whole, these placebo tests provide strong support to the robustness of our main results on fiscal discipline.

1.6.3 Political economy of fiscal adjustment

In this section, we exploit our research design to shed light on the political economy of deficit. Evaluating the differential response of different politicians and voters to an exogenous (albeit local) variation in fiscal rules, in fact, can identify important determinants of politically motivated deficits and provide new evidence about the costs and benefits of fiscal restraints. We start by looking at three political factors. First, we consider whether there are more than two parties in the city council—which must approve the budget proposed by the mayor—to capture political fragmentation and potential common pool problems. Second, we consider whether mayors face a binding term limit or not, because mayors in their second term have no reelection incentives and no personal stake in the city’s budget for the following years. Third, we consider the age profile of citizens in our municipalities. We then relate our findings to models that formalize the trade-off between reduced flexibility of fiscal policy and increased discipline on politically motivated deficit. First, we consider the effort of the mayor in providing the public good promised to voters in the provisional budget. Second, we look at cities characterized by historical deficit versus those characterized by a cyclical one.

Results for the first two variables are reported in Table 1.7, where we implement the baseline diff-in-disc estimations in split samples: (i) two parties in the city council (the median value) vs. more than two parties; (ii) binding vs. non-binding term limit. For each heterogeneity exercise, we report the diff-in-disc estimates in the two (split) subsamples, the difference between the two point estimates, as well as the Wald test p-value indicating whether this difference is statistically different from zero. We are aware that, while the point estimates in each subsample consistently identify local average treatment effects, the causal interpretation of their difference rests on an additional conditional

independence assumption. This is why we also report a second Wald test p-value (with covariates) indicating whether this difference is robust to a specification including a full set of interactions with covariates at the municipality level (namely, the average taxable income; mayor's years of schooling; and whether the municipality is in the North of the country). If also this test is statistically significant, it means that the differential impact of relaxing fiscal rules across our heterogeneity dimensions is not driven by other confounding city characteristics.

First, we focus on political fragmentation. Political fragmentation generally arises when several agents have an active role in the allocation of the budget, each with its own constituency to please, and each with some weight in the final decision. There are two key determinants that affect how much a policy maker internalizes the costs of the demanded share of the budget: the number of decision makers participating in the bargaining process and the institutional rules determining the aggregation of preferences. Most empirical studies focus on the first determinant because of a lack of reliable proxies for the rules that determine the budget allocation across countries. We also follow this previous literature by focusing on the first determinant (see Kontopoulos and Perotti, 1999). However, one advantage of our setting is that we can safely assume that the rules that determine the allocation of the budget are constant around our threshold. The estimation results reported in Table 1.7 show that only municipalities with high political fragmentation react to the relaxation of the DSP, and this result is robust to controlling for covariates. This result is consistent with models that explain deficit in terms of political fragmentation or dynamic common pool (see Persson and Tabellini, 2000) and also with the cross-country evidence that coalition governments lead to higher deficits (see Roubini and Sachs, 1989; Kontopoulos and Perotti, 1999).

Second, we focus on term limit, exploiting the fact that Italian mayors face a two-term limit.³⁵ Theoretical models suggest that the expectation of a future election can

³⁵It should be noted that: (i) municipalities do not vote at the same time, and (ii) the DSP was independent of local politics because it followed agreements between the European Union and its member countries.

Table 1.7: *The political economy of deficit bias, part I*

	LLR $h = 500$	LLR $h = 750$	Spline poly 3 rd	Spline poly 4 th
With two parties or less:				
Deficit	0.720 (9.511)	2.753 (7.650)	3.278 (10.369)	2.993 (12.379)
Obs.	1,187	1,721	3,584	3,584
With more than two parties:				
Deficit	44.096*** (15.429)	42.958*** (12.082)	50.869*** (19.446)	69.627** (27.413)
Obs.	893	1,347	2,716	2,716
<i>Difference between the two subsamples</i>	43.376	40.205	47.591	66.634
<i>Wald test p-value without covariates</i>	0.034	0.015	0.022	0.044
<i>Wald test p-value with covariates</i>	0.037	0.028	0.028	0.049
With binding term limit:				
Deficit	0.967 (8.463)	7.861 (6.957)	4.017 (9.599)	5.598 (12.083)
Obs.	920	1,375	2,780	2,780
Without binding term limit:				
Deficit	29.531** (13.120)	27.189*** (9.428)	33.047** (14.471)	36.640* (20.332)
Obs.	1,160	1,693	3,520	3,520
<i>Difference between the two subsamples</i>	28.564	19.328	29.030	31.042
<i>Wald test p-value without covariates</i>	0.071	0.088	0.090	0.189
<i>Wald test p-value with covariates</i>	0.097	0.064	0.096	0.227

Notes. Municipalities between 3,500 and 7,000 inhabitants; budget years between 1999 and 2004. Diff-in-disc estimates of the impact of relaxing fiscal rules on fiscal discipline below 5,000 after 2001 in different subsamples (that is, above vs. below median number of parties; binding vs. non-binding term limit). Estimation methods: Local Linear Regression (LLR) with bandwidth $h = 500$ or $h = 750$, as in equation (1.5); spline polynomial approximation with 3rd-order or 4th-order polynomial, as in equation (1.6). The *Wald test p-value without covariates* evaluates whether the estimates are statistically different in the two subsamples. The *Wald test p-value with covariates* evaluates whether the estimates are statistically different in the two subsamples also controlling for a full set of interactions between the above specifications and appropriate covariates, such as: average taxable income; mayor's years of schooling; and whether the municipality is in the North. All variables are per capita and in 2009 Euros. Robust standard errors clustered at the municipality level are in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

affect policies because politicians who care to run again for office must please the voters sufficiently often to merit reelection (Barro, 1973; Banks and Sundaram, 1998). We find that the fiscal (de)stabilization induced by the relaxation of the fiscal restraints is driven by mayors without a binding term limit, although this result becomes borderline insignificant in some specifications that control for covariates. As mayors without term limit face both stronger reelection concerns and a higher expected probability that they (or their party) will remain in power, the above result provides more support for models linking deficit to reelection incentives (see Aghion and Bolton, 1990) or to politicians' pandering to voters (see Maskin and Tirole, 2004), rather than models viewing deficit as a way to tie the hands of future governments with different political preferences (see Alesina and Tabellini, 1990; Persson and Svensson, 1989; Tabellini and Alesina, 1990). Unfortunately, we are not able to provide further empirical evidence on strategic voting models because of the lack of a clear expected reelection probability outcome in our data.³⁶ We are also not able to rule out alternative channels that can rationalize our result on term limit, such as the possibility that political experience per se has an effect on how mayors react to the relaxation of fiscal rules, even if it is encouraging that some of the results survive to controlling for the mayor's characteristics.³⁷

In Table 1.8, which has the same structure of Table 1.7, we report one additional heterogeneity result along city characteristics. In particular, we implement the baseline diff-in-disc estimations in two separate subsamples: cities with a higher (i.e., above-median) fraction of young cohorts vs. the rest of the cities. Consistently with the model of Song, Storesletten, and Zilibotti (2012), deficit increases after the relaxation of the rule only in cities with a larger proportion of young citizens. Song, Storesletten, and Zilibotti (2012) propose a dynamic politico-economic theory of fiscal policy for small open economies, and we view this model as particularly relevant for our setting because small cities are a reasonably good

³⁶See Petterson-Lidbom (2001) for an empirical evaluation of strategic voting models.

³⁷Given that we can control for the selection of mayors (years of schooling), our findings provide additional support to the literature that focuses on the effect of term limit on political accountability and in-office performance (see Besley and Case, 1995; List and Sturm, 2006).

approximation of small open economies. The main intuition of their model is that younger citizens impose a disciplining effect on fiscal policy, because they internalize the future costs of a loose fiscal policy today. Both of the above predictions are borne out by our empirical findings.³⁸

Table 1.8: *The political economy of deficit bias, part II*

	LLR $h = 500$	LLR $h = 750$	Spline poly 3 rd	Spline poly 4 th
Young cohorts above median:				
Deficit	0.887	8.110	7.094	4.151
	(10.509)	(7.753)	(11.067)	(14.635)
Obs.	1,059	1,561	3,224	3,244
Young cohorts below median:				
Deficit	35.523***	31.244***	36.913***	49.481***
	(12.010)	(9.390)	(14.255)	(18.795)
Obs.	1,021	1,507	3,076	3,076
<i>Difference between the two subsamples</i>	34.636	23.134	29.819	45.330
<i>Wald test p-value without covariates</i>	0.024	0.050	0.082	0.041
<i>Wald test p-value with covariates</i>	0.052	0.068	0.156	0.048

Notes. Municipalities between 3,500 and 7,000 inhabitants; budget years between 1999 and 2004. Diff-in-disc estimates of the impact of relaxing fiscal rules on fiscal discipline below 5,000 after 2001 in different subsamples (that is, above vs. below median percentage of young cohorts). Estimation methods: Local Linear Regression (LLR) with bandwidth $h = 500$ or $h = 750$, as in equation (1.5); spline polynomial approximation with 3rd-order or 4th-order polynomial, as in equation (1.6). The *Wald test p-value without covariates* evaluates whether the estimates are statistically different in the two subsamples. The *Wald test p-value with covariates* evaluates whether the estimates are statistically different in the two subsamples also controlling for a full set of interactions between the above specifications and appropriate covariates, such as: average taxable income; mayor's years of schooling; and whether the municipality is in the North. All variables are per capita and in 2009 Euros. Robust standard errors clustered at the municipality level are in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Finally, we move to relating our results to existing models that formalize the costs and benefits of restraining fiscal policy. Besley (2007) considers fiscal rules that increase the cost of issuing public debt in an environment characterized by two types of politicians, the ones who care about smoothing business cycle fluctuations and those who care about reelection.

³⁸Another testable prediction of their model is that deficit bias should be higher for cities where the mayor is affiliated with right-wing parties. While most of the mayors in our sample are not affiliated with parties that can be clearly mapped to the ideological spectrum, we also find evidence that is consistent with this prediction in the (small) subset of mayors affiliated to a political party (results available upon request).

Rules that are not overly restrictive will bind more on bad politicians and increase voters' welfare. In Table 1.9, we split the sample based on the median of the speed of public good provision (measured as the ratio of paid outlays over the total outlays committed in the provisional budget). Our indicator of the speed of public good provision is calculated as a mayor-specific measure, averaged across all the years each mayor is in office. The results show that the increase in deficit bias arises only for mayors that systematically under-provide the public good promised to voters in the provisional budget. This evidence can be interpreted as consistent with a potential negative effect for voters following the relaxation of the rule, only to the extent to which politicians who consistently over-promise public goods are those who care more about reelection rather than social welfare.³⁹

In the second panel of Table 1.9, we look at whether the fiscal (de)stabilization following the relaxation of the DSP is driven by cities with an historical rather than a cyclical deficit in the pre-treatment period, by going as back in time as possible with our data.⁴⁰ Specifically, we classify as municipalities with "historical deficit" those characterized by deficit in all the pre-treatment years, or in all the pre-treatment years but one. One important limitation of this last heterogeneity analysis is that we have limited power for measuring the historical propensity to accumulate debt as we can only rely on four pre-treatment years.⁴¹

Battaglini and Coate (2008) predict that the costs of restricting fiscal rules will be higher when governments are characterized by cyclical, rather than persistent, deficit. We find that the cities who increase deficit are not those characterized by cyclical deficit, and, if anything, those with more persistent deficit are those who accumulate more debt when unconstrained, although this last result is not statistically significant.⁴²

³⁹Rogoff (1990) argues that electoral incentives might distort fiscal policy because of the distorted incentives to over-provide public good when it is more salient for voters.

⁴⁰Note that we performed an additional robustness check for the five heterogeneity dimensions that we study in this section. Specifically, we checked whether these variables are balanced at the 5,000 threshold. This is indeed what we find, suggesting that our strategy of splitting the sample is keeping the sample balanced in the neighborhood of 5,000 inhabitants as well (results available upon request).

⁴¹We were able to construct an internally consistent sample of administrative data from the Italian Ministry of the Interior only starting from 1997.

⁴²As a final robustness check on our heterogeneity analysis, we repeat the falsification test in 1999 for all

Table 1.9: Fiscal restraints and budget management

	LLR $h = 500$	LLR $h = 750$	Spline poly 3 rd	Spline poly 4 th
Speed of public good provision above median:				
Deficit	-0.851 (11.291)	7.946 (7.880)	4.433 (11.602)	3.717 (16.051)
Obs.	1,022	1,492	3,153	3,153
Speed of public good provision below median:				
Deficit	35.082*** (11.995)	29.395*** (9.498)	38.184** (14.811)	47.452** (20.275)
Obs.	1,058	1,576	3,147	3,147
<i>Difference between the two subsamples</i>	35.932	21.449	33.751	43.735
<i>Wald test p-value without covariates</i>	0.025	0.075	0.065	0.086
<i>Wald test p-value with covariates</i>	0.029	0.109	0.063	0.057
Historical deficit:				
Deficit	25.023*** (6.599)	19.831*** (6.300)	26.365*** (7.913)	26.192*** (9.260)
Obs.	1,536	2,256	4,648	4,648
No historical deficit:				
Deficit	6.481 (17.967)	4.279 (13.156)	10.843 (20.978)	10.197 (30.104)
Obs.	912	1,332	2,712	2,712
<i>Difference between the two subsamples</i>	18.542	15.552	15.522	15.995
<i>Wald test p-value without covariates</i>	0.332	0.286	0.488	0.611
<i>Wald test p-value with covariates</i>	0.264	0.342	0.557	0.453

Notes. Municipalities between 3,500 and 7,000 inhabitants; between 1999 and 2004. Diff-in-disc estimates of the impact of relaxing fiscal rules on fiscal discipline below 5,000 after 2001 in different subsamples (that is, above vs. below median speed of public good provision; historical vs. no historical deficit). Municipalities with historical deficit are those characterized by deficit in all the pre-treatment years since 1997, or in all the pre-treatment years but one. Estimation methods: Local Linear Regression (LLR) with bandwidth $h = 500$ or $h = 750$, as in equation (1.5); spline polynomial approximation with 3rd-order or 4th-order polynomial, as in equation (1.6). The *Wald test p-value without covariates* evaluates whether the estimates are statistically different in the two subsamples. The *Wald test p-value with covariates* evaluates whether the estimates are statistically different in the two subsamples also controlling for a full set of interactions between the above specifications and appropriate covariates, such as: average taxable income; mayor's years of schooling; and whether the municipality is in the North. All variables are per capita and in 2009 Euros. Robust standard errors clustered at the municipality level are in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

the above dimensions but historical deficit (which is constructed using data on the pre-treatment period and thus cannot be used in that context). Specifically, in the Appendix Tables A5, A6, and A7 we implement the above heterogeneity diff-in-disc estimations in the interval 1997–2000, using 1999–2000 as the post-treatment period and 1997–1998 as the pre-treatment period. The fact that no effect and no difference are ever statistically significant means that municipalities around the threshold in different heterogeneity subsamples are not on differential trends before 2001. In other words, the DSP did not bind differently across those subsamples.

On the whole, the results discussed in this section suggest that fiscal restraints are more likely to bind for cities characterized by political failures, and that political economy factors play a first-order role in the process of fiscal adjustment.

1.7 Conclusion

Limiting the increase of public debt is a key policy issue in most economies. Fiscal rules are usually considered one of the potential solutions to public debt growth. In this paper we rely on a novel quasi-experimental design to show that fiscal rules enforced by a national government can be effective in causing a reduction of the accumulation of debt by local governments. Additionally, we are able to investigate the composition of fiscal adjustment and we show that unconstrained cities have lower tax rates and lower revenues following the policy change. We then link our results to existing theories of fiscal adjustment to provide new evidence about the costs and benefits of restraining fiscal policy. We show that deficit bias arises only where many parties seat in the city council, where mayors can run for reelection, where there is a smaller fraction of young citizens, where mayors systematically underprovide the promised public good, and in cities characterized by historical deficit. These results suggest that fiscal restraints can be more effective when political distortions are larger.

We are aware that the enhanced internal validity of our evaluation design comes at the price of lower external validity, as it is always the case in (local) econometric strategies based on policy discontinuities. However, our falsification test shows that mayors who are selected through a different mechanism (that is, a different salary) react in the same way to the introduction of fiscal rules. This suggests that our results can potentially apply to different institutional settings.

Our results raise a number of questions for further research. First, we show that fiscal rules, when accompanied by a proper enforcement mechanism, can be effective also in regulatory environments characterized by serious commitment issues such as the Italian case. Hence, fiscal rules might be useful in far more cases than those suggested by the

conventional wisdom, and the optimal design of fiscal rules should take into account political incentives in the enforcement of the rules. Second, our results on the composition of fiscal adjustment suggest that stabilizing fiscal policy through revenues rather than expenditures might not be politically equivalent. A rapidly growing literature in public economics have shown how the salience of tax changes affects behavioral responses (see Chetty, Looney and Kroft, 2009). It is an exciting direction for future research to investigate whether public good provision is subject to similar issues, and whether policy makers can exploit voters' behavioral biases in their favor.⁴³

Our last set of empirical results implies that political incentives drive local government responses to exogenously imposed fiscal restraints. Since restricting tax smoothing is the main welfare cost of fiscal restraints identified by the macroeconomic literature, and since mechanisms for smoothing business cycle fluctuations, such as unemployment insurance, are often administered at the national level (as suggested by Gavin and Perotti, 1997), our results suggest that fiscal rules imposed on subnational governments might have limited welfare costs and significant benefits.

⁴³A first attempt along these lines is Bisin, Lizzeri, and Yariv (2011).

Chapter 2

Ghost-House Busters: The Electoral Response to a Large Anti Tax Evasion Program¹

2.1 Introduction

Government ability to enforce tax collection efficiently is one of the fundamental components of state capacity and, in turn, an important driver of historical economic development. Tax evasion generates both large losses in government revenues and large distortions.² The literature (e.g., Slemrod (2007); Besley and Persson (2012)) describes three main determinants of tax compliance: enforcement technology, political incentives, and cultural norms. This paper illustrates the interaction among these three factors. We estimate the electoral returns — the change in reelection likelihood — that local policymakers obtain from a nationwide anti tax evasion policy in Italy, based on an innovation in tax-payers monitoring technology. In

¹Co-authored with Lorenzo Casaburi.

²Slemrod (2007) estimates overall noncompliance in the United States at 14 percent. Estimates from other developed countries deliver similar figures (for Italy, Marino and Zizza (2008)). In developing countries, where the share of the informal economy is typically larger, the figures are much higher (Gordon and Li (2009); Schneider, Buehn, and Enste (2010)).

addition, we study how these electoral returns depend on underlying social preferences for tax compliance and on the local government efficiency in public good provision. The paper provides, to the best of our knowledge, the first empirical evidence on voters' responses to anti tax evasion policies.

Measures to reduce tax evasion generate a conflict between voters. They hurt tax evaders, typically a minority of voters, while the majority of the population is likely to benefit from additional government expenditures, lower tax rates, or even directly from the punishment of former shirkers.³ However, the magnitude of the individual costs tax evaders incur is potentially higher than the individual benefits non evaders derive. Anti tax evasion policies are thus canonical examples of policies that have an asymmetry in the concentration of costs and benefits (Tullock (1959); Olson (1965)). Depending on which of the two types of voters is more likely to change its voting behavior in response to changes in enforcement, fighting tax evasion might either benefit or harm politicians who seek reelection. The sign of this impact is ex ante ambiguous and, therefore, an empirical question.

In 2007, the Italian government instituted a nationwide anti tax evasion policy, the *Ghost Buildings* program. The program identified ghost buildings — properties not included in the land registry and thus hidden from tax authorities⁴ — by overlaying aerial photographs and digital land registry maps.⁵ The intervention detected more than two million land registry parcels with ghost buildings.⁶ Following the completion of the mapping exercise, the government commenced a large registration program that targeted the identified ghost buildings. While the central government began the program and coordinated registration activities, municipality administrations circulated information about the program, collab-

³For experimental evidence on this channel, see Carpenter et al. (2009); Casari and Luini (2009); Ouss and Peysakhovich (2012).

⁴The value of registered buildings enters the tax base for personal income tax and property taxes, among other taxes.

⁵Other countries, such as Greece and Rwanda, have recently implemented policies using similar technologies.

⁶The unit of the Italian land registry maps is the parcel (*parcella*), which is defined as a portion of land belonging to a given physical or legal person. In the case where the land is shared across several owners, the parcel is split into several sub-parcels (*subalterni*).

orated with follow-up inspections, and enforced payment of overdue local taxes. Media reports highlight both the importance of local administrations in the registration process⁷ and the heterogeneity in their actions in response to the program.⁸

The policy induced a large shift in tax enforcement, the intensity of which varied significantly across towns. In towns with a higher prevalence of detected ghost buildings, the program had larger scope to affect the level of building registration. We use a measure of *Ghost Building Intensity* — the ratio of the number of land registry parcels with ghost buildings identified by the program to the total number of land registry parcels in the town — to proxy for the scope of the program. Using a difference-in-differences approach, we test the impact of the anti evasion policy on local incumbent reelection by exploiting variations across municipalities in this intensity.⁹ This strategy, under plausible assumptions, isolates the causal effect on electoral outcomes of the policy scope to increase enforcement from other mayor or voter characteristics that might have affected the actual levels of ghost building registration in the town.¹⁰

In local elections occurring after the beginning of the program, an increase of one standard deviation in the ghost building intensity raises the likelihood of reelection of the local incumbent relative to pre-program elections by approximately 2.5 percentage points, about 5.5% of the average reelection rate. Higher town-level ghost building intensity also lowers several measures of competitiveness of local elections. In particular, it reduces the number of candidates running for election, increases the margin of victory for the winner, and reduces the likelihood of a runoff. Guiding our empirical models is a retrospective voting framework of political agency. Such a theoretical framework helps us predict how a change in tax enforcement can impact voter choices and which factors affect this response.

⁷For example, Dell'Oste and Trovati (2011).

⁸Among many others, Bernardini (2011) and Barca (2008) discuss the particular way in which the city of Montecatini and some cities in the Reggio Emilia province implemented the program, respectively.

⁹In a recent contribution, Mian and Sufi (2012) adopted a similar empirical approach to study the effects of the fiscal stimulus in the US.

¹⁰We verify that the assumptions required by the identification strategy hold (i.e., no contemporary differential changes and no differential pre-trends in the outcome variable by treatment intensity).

Additional analysis of the actual building registrations induced by the program complements the reduced form analysis described above. For a given town-level program scope, a higher registration rate of ghost buildings under the incumbent local administration (i.e., the share of ghost buildings that gets registered prior to the local election date) has a positive effect on the likelihood of reelection. The result is robust to the inclusion of mayors' characteristics as controls and to an instrumental variable approach, based on the time elapsed between the program start date and the town election date.

We provide evidence for two channels that could drive the observed electoral response. First, towns where the government is more efficient in delivering public goods show a larger electoral response to the program. We also verify that towns with higher ghost building intensity experienced a differential increase in local government expenditures following the program inception. Second, using survey data on the self-reported tolerance for tax evasion among voters, we show that the program's positive impact on incumbent reelection is significantly higher in areas with lower tolerance for tax evasion. Finally, the empirical findings are inconsistent with two potential alternative interpretations on the impact of the program on voter support for incumbents. In the first, the program changes voter behavior by providing information on the existing stock of ghost buildings. In the second, it gives an electoral rent to the incumbent by giving her the option to *not* register identified ghost buildings.

Our approach can potentially be applied in different settings to study the political feasibility of upgrading tax administrations around the world using new electronic data, cross-checking technologies, and other monitoring devices (Bird and Zolt (2008)). Additionally, our analysis points at complementarity between technological innovations in tax enforcement and political incentives. When exposed to a reduction in monitoring costs, politicians exploit the new technologies and experience political gains. These findings have a direct bearing on the political feasibility of upgrading tax administrations around the world using new electronic data, cross-checking technologies, and other monitoring devices (Bird and Zolt (2008)). In addition, our study provides evidence that the underlying tax

culture shapes the political incentives for tax enforcement and the political returns to these innovations (Torgler (2007); Rothstein (2000)). We discuss several policy implications arising from these findings. Finally, access to town-level nationwide administrative data from the program allows us to provide evidence on two additional fronts. First, we study the correlates of tax evasion at the town level. We find that geographical features, such as town size, are important determinants of tax evasion, consistent with Saiz (2010), and that social capital is negatively correlated with tax evasion (Putnam (2001)). Second, we document that mayor characteristics, such as education, gender, and age, do affect the extent to which the Ghost Buildings program increased tax enforcement (consistent with Alesina (1988); Besley and Coate (1997); Besley, Montalvo and Reynal-Querol (2012)).

This paper relates to several strands of literature. First, a recent set of studies uses microdata to shed light on enforcement technologies such as third-party reporting (Slemrod, Blumenthal, and Christian (2001); Saez (2010); Kleven et al (2011); Chetty, Friedman, and Saez (2012)), paper trails (Pomeranz (2012); Kumler, Verhoogen, and Frías (2011)), cross-checking (Carrillo, Pomeranz, and Singhal (2012)), and targeted auditing strategies (Almunia and Lopez-Rodriguez (2012); Aparicio (2012)).¹¹ By studying how technology-driven enforcement policies affect policymakers, we bridge this work with the one estimating the political returns to fiscal policies (Brender and Drazen (2008); Alesina, Carloni, and Lecce (2011)). In addition, by delving into the relation between incentives of political agents and tax evasion, our paper is related to Artavanis, Morse, and Tsoutsoura (2012), who find that tax evasion is higher in industries supported by parliamentarians. Finally, our results provide support to the existing literature that highlights the role of culture and social norms as determinants of tax evasion, either via cross-country analysis (Torgler (2003); Slemrod (2003)) or lab experiments (Spicer and Becker (1980); Alm, Jackson, and McKee (1992)).

The remainder of the paper is organized as it follows. Section 2.2 describes the Ghost Buildings program. Section 2.3 presents a simple framework that guides our empirical analysis. Section 2.4 describes the data and presents descriptive evidence. Section 2.5

¹¹For a review of the literature, see Andreoni, Erard, and Feinstein (1998), Slemrod and Yitzhakil (2002).

lays out our empirical strategy to estimate the electoral response to the policy. Section 2.6 presents the results. Section 2.7 concludes.

2.2 The Ghost Buildings Program

The value of the buildings registered in the land registry enters the tax base for several national and local taxes, including “ICI”, the local property tax, “IRPEF”, the personal income tax,¹² and the local waste management tax. Italian legislation¹³ requires that owners register new buildings at the local office of the *Agenzia del Territorio*, the agency managing the land registry, within thirty days after their completion.¹⁴

In 2006, the national government approved new anti tax evasion legislation, the Ghost Buildings program,¹⁵ aimed at detecting buildings not registered in the land registry maps.¹⁶ The *Agenzia del Territorio*, the national agency managing the land registry, coordinated the effort. The *Agenzia del Territorio* first juxtaposed the land and building registry maps to obtain the “Official Building Map”. It subsequently compiled high-resolution (50cm) aerial photographs of the entire country in order to identify the ghost buildings. Figures 1a-1c summarize the steps of the identification. First, the aerial photograph for a particular location was created (Figure 1a). Second, the pictures were matched with the official building map for the corresponding area (Figure 1b). Finally, the ghost buildings were identified (Figure 1c)¹⁷. Ghost buildings include commercial, industrial, and residential

¹²“IRPEF” includes the inferred opportunity cost of living in the house, with both a local and a national component.

¹³*Legge 9 Marzo 2006 n.80 - Art. 34-quinquies.*

¹⁴All buildings in Italy need a building permit before their construction starts. Obtaining a building permit makes the building part of the City Plan. The process of obtaining building permits is administered independently from the registration in the land registry maps. Buildings not in the City Plan are required to be demolished.

¹⁵*Legge 24 novembre 2006, n. 286* subsequently modified by *Legge 30 Luglio 2010, n. 122.*

¹⁶The exercise did not cover one of the semi-autonomous regions, Trentino Alto-Adige, because in that region land registry maps are autonomously administered. The region contains less than two percent of the total population of Italy.

¹⁷According to the Law *Decreto Ministero delle Finanze 2 gennaio 1998, n.28.Art. 3* the following buildings do

stand-alone buildings, and also substantial extensions of previously registered buildings that should have been reported to the land registry.

Through this process, the *Agenzia del Territorio* identified approximately two million land registry parcels with unregistered buildings. Beginning in August 2007, the *Agenzia del Territorio* started to publish parcel-level data on unregistered properties in the *Gazzetta Ufficiale*, the official bulletin promulgating Italian laws and decrees, in order to induce registrations of the ghost buildings. Within three years, it coded detailed information on the number of ghost buildings in the universe of Italian municipalities (with the exception of Trentino Alto-Adige). The order of publication relied on the availability of digitized land registry maps at the time when the program started. The *Agenzia del Territorio* had 60% of the land registry maps of the Italian territory in digitized form before the Ghost Buildings program was approved. After 2006, the *Agenzia del Territorio* began digitizing the remaining land registry maps, proceeding by province (i.e., they simultaneously coded municipalities in the same province). It completed the identification exercise by the end of 2010.¹⁸

According to the initial legislation, owners could register the detected ghost building with the land registry by April 30, 2011.¹⁹ Widespread media campaigns and local administration efforts helped the program achieve high registration rates. In particular, local administrators a) disseminated information about targeted parcels; b) collaborated on follow-up building inspections; c) proceeded with the collection of overdue local taxes up to five years before the program began; and d) verified the conformity of ghost buildings to the city plan and local zoning restrictions. Local administrations received a large share of the additional tax revenues generated by the program. Owners of ghost buildings that registered prior to the April 2011 deadline were required to pay overdue taxes dating back to 2007, or to

not increase the tax base of their owners and thus are not subject to registration requirements: (i) buildings that are not completed (ii) buildings particularly degraded (iii) solar collectors (iv) greenhouses (v) henhouses or others reserved for animals.

¹⁸Publication on the *Gazzetta Ufficiale* occurred in the following waves: August 2007, October 2007, December 2007, December 2008, December 2009, December 2010.

¹⁹This was the result of two previous deadlines of ninety days and seven months since the publication on the *Gazzetta Ufficiale*.

Figure 2.1: *The Ghost Building Identification Process*

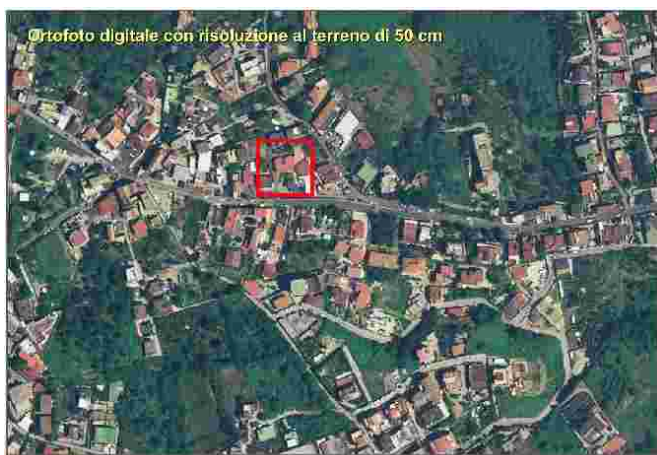


Figure 1A: Aerial Picture



Figure 1B: Digital Land Registry Map

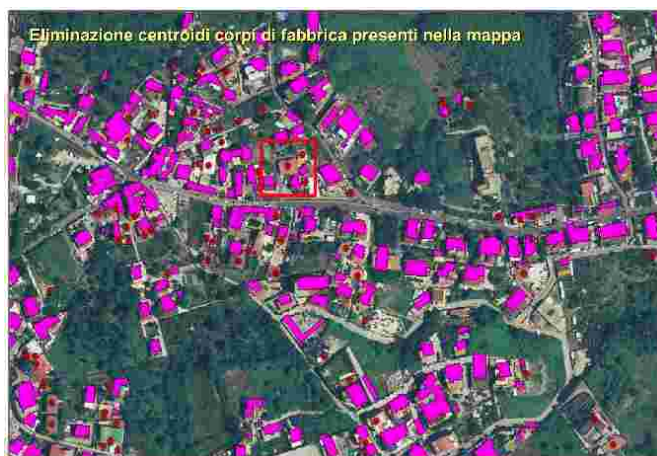


Figure 1C: Overlay

Source: Agenzia del Territorio

the construction date for post-2007 cases, and to pay penalties for delayed payments.²⁰ After April 2011, the *Agenzia del Territorio*, with the support of local administrations and local contractors, proceeded with follow-up inspections²¹ to impute the tax base for the remaining unregistered buildings.²² Additional penalties and a fee for the extra inspection were assessed on owners of buildings for which the *Agenzia del Territorio* imputed the tax base after April 2011.

The *Agenzia del Territorio* published the detailed economic impact of the program for the year 2011. The program led to a substantial wave of registrations. According to the administrative data, roughly 40 percent of the ghost buildings were registered as of 30th of December 2011. According to the figures provided, the program increased total tax revenues by 472 million euros in that year.²³ We estimate that approximately 65 percent of those revenues come from local taxes. We run a “back-of-the-envelope” calculation using figures on the number of land parcels with ghost buildings, the registration rates, and the total additional tax revenues from the program. A one standard deviation increase in ghost buildings targeted by the program increased the tax revenues by around 3.5% of the median value. Using the same information, we find that, on average, the owner of a registered ghost building now faces an additional yearly tax burden of approximately 528 Euros.

2.3 Tax Enforcement and Retrospective Voting: A Conceptual Framework

This section outlines the impact of a change in tax enforcement on voters’ electoral choices. We provide a simple framework based on modeling of retrospective voting (Barro (1973);

²⁰Penalties were determined by *Legge 29 Dicembre 1990, n. 408*, subsequently modified by *Decreto Legislativo. 18 Dicembre 1997*.

²¹To increase incentives for the local administrations further, an additional bonus was introduced in 2011 for each registered ghost building.

²²*Decreto Legge 79/2010, art. 10, 11*.

²³This figure does not include payment for overdue taxes from previous years.

Ferejohn (1986)) and of tax evasion (Allingham and Sandmo (1972)). The main intuition of retrospective voting models is that citizens examine whether their welfare has increased under a politician's office, and vote accordingly.²⁴ While the specific application to tax evasion is novel, the discussion in this section heavily relies on the intuition from existing models.

The economy is populated by a unit mass of voters and by politicians. Voters are heterogeneous in their ability to evade. This ability could be a function of occupation type (employed vs. self-dependent) or evasion costs (economic and psychological). We consider a simple case with two fixed types of voters: evaders and non evaders. Evaders pay taxes only if enforcement occurs, while non-evaders always pay taxes.²⁵ The population share of evaders is λ . Enforcement of tax collection for each evader occurs with probability p . Enforcement draws are independent across evaders, and thus p is the share of evaders for which enforcement occurs. This is assumed to be a function of the politician type (a) and of an idiosyncratic component (v), whose distributions are $G(a)$ and $G(v)$, respectively. Voters do not observe the two components and are uncertain over the politician type (Banks and Sundaram (1998)). They use previous realizations to form expectations \hat{a} and \hat{p} (in the spirit of Holmstrom, 1982).

We assume an exogenous income level (normalized to 1) and tax rate (τ), constant across the population. Voters derive utility from disposable income and from the overall level of enforcement, for instance through increased public good provision and deficit reduction. This implies that enforcement has two effects on evaders' utility, which go in opposite directions. First, enforcement decreases the disposable income for evaders. However, cracking down on tax evasion increases the size of government, which benefits all

²⁴This implies that incumbent's reelection is the main outcome to analyze when testing the empirical predictions of retrospective voting models. Retrospective voting models have received considerable empirical support in the context of government corruption or fiscal stabilization (Brender, 2003; Besley and Pratt, 2006; Ferraz and Finan, 2008; Nannicini, Stella, Tabellini and Troiano, 2012).

²⁵For simplicity, we ignore the extra fines evaders pay when audited and, thus, the optimal individual evasion level they choose.

citizens, including evaders. The expected utility for evaders, V_E is defined as:

$$V_E(\hat{p}) = \hat{p}(U(1 - \tau)) + (1 - \hat{p})U(1) + \hat{p}W_E(\lambda, e), \quad (2.1)$$

where we highlight that V_E depends on the expected level of enforcement, \hat{p} . In Equation 2.1, $U(\cdot)$ is the monetary utility from disposable income and $W_E(\cdot)$ is the utility from tax collection enforcement.²⁶ W_E is increasing in λ , the share of evaders in the population, and e , the government efficiency in using tax revenues to produce public good.

We allow non-evaders to obtain an additional non-monetary benefit from enforcement. One example is the case where, because of fairness concerns, non-evaders derive direct utility from the enforcement of evaders' tax payments, independently from their monetary returns. Thus, the expected utility function for the non-evaders is:

$$V_N(\hat{p}) = \hat{p}W_N(\lambda, e, n) + U(1 - \tau) \quad (2.2)$$

We notice that, in addition to λ and e , V_N also depends on n , a shifter that affects the non-monetary benefits from increases in enforcement. For instance, n captures the extent to which voters are averse to tax evasion ("tax culture"). In the model, we abstract from the utility arising from government services financed by the tax payments of the non-evaders since that does not depend on \hat{p} , the core variable of interest for our argument.

We now consider the voters' choice between an incumbent and a contender. We adopt a standard probabilistic voting approach (Lindbeck and Weibull, 1987). In the text below, \hat{a} and \hat{p} denote the voters' beliefs about incumbent type and enforcement, respectively. On the other hand \bar{a} and \bar{p} capture the expectations about the contender. In deciding whether to reelect the incumbent, the two groups of voters compare the utility under the expected incumbent's type with an average opponent. Voter i in group $j = \{E, N\}$ will reelect the incumbent if $V_j(\hat{p}) > V_j(\bar{p}) + \epsilon_{ij} + \delta$. The parameter ϵ_{ij} is an individual ideological bias

²⁶In order to simplify the presentation, we assume that the utility from enforcement is proportional to the expected level of enforcement.

toward the contender, distributed uniformly over $[-\frac{1}{2\phi^I}, \frac{1}{2\phi^I}]$.²⁷ The parameter δ measures the average popularity of the contender in the population and is distributed uniformly over $U[-\frac{1}{2}, \frac{1}{2}]$.

Under the above assumptions, the ex-ante incumbent reelection probability (i.e., before the realization of δ) is:

$$\pi = (\hat{p} - \bar{p}) [\lambda\phi_E(-U(1) + U(1 - \tau) + W_E) + (1 - \lambda)\phi_N W_N] \quad (2.3)$$

The following equation presents the electoral impact of an increase in the expected enforcement level under the incumbent, \hat{p} :

$$\frac{\partial \pi}{\partial \hat{p}} = \lambda\phi_E(-U(1) + U(1 - \tau) + W_E) + (1 - \lambda)\phi_N W_N \quad (2.4)$$

The first term of the outer sum represents the net electoral gains coming from evaders voting. These will be negative whenever the utility cost of the expected loss in disposable income, $U(1) - U(1 - \tau)$, more than offsets the benefits from enforcement, W_E . The second term is the electoral gain from non-evaders (always positive). This duality is consistent with the discussion in Section 2.1: an increase in the perception of the enforcement type of the incumbent has ambiguous effects. The change generates a conflict across voters and the model parameters determine which channel prevails.

In addition, the model delivers intuitive comparative statics on the heterogeneity of the electoral impact arising from an increase in expected enforcement under the incumbent. Intuitively, both governmental efficiency in public good provision and the intensity of non-monetary benefits from the additional enforcement matter play a role. Specifically:

$$\frac{\partial^2 \pi}{\partial \hat{p} \partial e} = \lambda\phi_E \frac{\partial W_E}{\partial e} + (1 - \lambda)\phi_N \frac{\partial W_N}{\partial e} \quad (2.5)$$

and

$$\frac{\partial^2 \pi}{\partial \hat{p} \partial n} = (1 - \lambda)\phi_N \frac{\partial W_N}{\partial n}, \quad (2.6)$$

²⁷The parameters ϕ^E and ϕ^N should be interpreted as proxies for the responsiveness of voters in each group to tax evasion enforcement. They might reflect for example the fact that the political power of a group can change depending on its size or ability to self-organize (Olson(1965)).

which are both positive.

To summarize, the simple model predicts that an increase in the expected level of enforcement under the incumbent:

- i) Has an ambiguous impact on the incumbent reelection likelihood
- ii) Is larger when government is more efficient in public good provision
- iii) Is larger when the non-monetary returns from enforcement are larger

The Ghost Buildings program allows us to shed light on these predictions. The, program initiated by the central government, can be considered as a positive shock to enforcement. We argue that voters observe the increase in the building registration but have limited information about the specific “production function” of enforcement (i.e., information collected by the central government, local administration effort, and complementarity between the two sources) This in turn increases the belief voters hold about the local incumbent type, \hat{a} (and thus on \hat{p}), and, according to the model, generates an ambiguous effect on the incumbent reelection probability.

Crucially, this result relies on the assumption that voters have limited information about the details of the Ghost Buildings program. They observe the change in enforcement and still attribute a part of it to the incumbent, thus extracting signal on her type. Models with *rational but poorly informed voters* have received growing attention in the literature. They can provide theoretical support for the empirical findings that voters’ electoral choices respond to economic conditions (Wolfers (2009)), natural disasters (Cole, Healy, and Werker (2012)), corruption (Nannicini, Stella, Tabellini and Troiano, (2012)) and quasi-random targeted transfers (Manacorda, Miguel, and Vigorito (2011)). In addition, a recent wave of randomized experiments shows that information provision can significantly affect voter choices and political outcomes (for a review, see Pande (2011)). For the Ghost Buildings program, it is likely that inference about who exactly was causing the extra enforcement was difficult. Local administration efforts complemented the initial identification process. In addition, evidence from media reports and town bulletins suggest that at least some

mayors claimed a role throughout the program, including the initial stages of building identification through aerial pictures (Cavallaro (2011), Corriere della Citta' (2012), Gazzetta del Mezzogiorno (2012)).

Finally, we notice that it is also possible to predict an impact on support for the incumbent in an alternative model where voters perfectly observe the nature of the Ghost Buildings program (while they are still uncertain about the type of mayor). In this alternative setting, the program provides an opportunity for voters to extract a more precise signal about the incumbent type, as in Bubb (2008). This can in turn either benefit or hurt the incumbent, depending for instance on voter risk preferences (Quattrone and Tversky (1988)) or on the skewness of the distribution of incumbent types (Caselli et al, (2013)). In the rest of the paper, we do not aim to differentiate the two classes of models in the data analysis. Rather, the insight that the net voter response to an enforcement policy is theoretically ambiguous, which is common to both models, motivates our empirical investigation.

2.4 Data and Descriptive Evidence

2.4.1 Data

The main database for the analysis includes information on the number of parcels containing Ghost Buildings in each town. The aerial photographs detected more than two million such parcels. We target the population of 7,720 of the 8,092 Italian towns (*Comuni*) for which we can define the measure of ghost building intensity. Additionally, we obtain data on registered ghost buildings up to the deadline of April 30, 2011. In order to analyze the electoral response to ghost building registration, we construct a measure of registration imputable to the incumbent administration. Specifically, we multiply the registration rate by the ratio between a) the time elapsed between program start date and election date and b) the time elapsed between program start date and April 2011.²⁸

²⁸In one of our robustness checks, we also compute a second measure of registration imputable to the incumbent under the assumption of a constant growth rate of 50% in the registration levels over years.

We complement this information with data from the Italian Department of the Interior (*Ministero degli Interni*) which contains outcomes for the universe of municipal elections from 1993 to 2011.²⁹ In Figure 2.2, we plot the number of elections per year. Towns vote in different years, according to predetermined waves. We distinguish between elections before and after the beginning of the Ghost Buildings program. There are almost 5,200 municipalities for which we have data on an election that occurred after program inception (about 67% of the total number of towns targeted by the program). It is also important to discuss two institutional reforms that occurred in the time span of our sample. First, in 1993, the starting year for our election sample, Italian municipal politics were overhauled: a new electoral law changed the mayoral electoral system from party to individual ballot. It also introduced a two-term limit. Second, in 2000, the length of the mayoral term was extended from four to five years.

In addition to the core data, we collect geographic and socio-economic data at the municipality level from the Italian National Statistical Office. Finally, we use two additional data sources to test the channels driving the electoral response: town-level government expenses (from the *Ministero degli Interni*) and a region-level standardized score to the question “Do you justify tax cheating?” from the *European Values Study*.

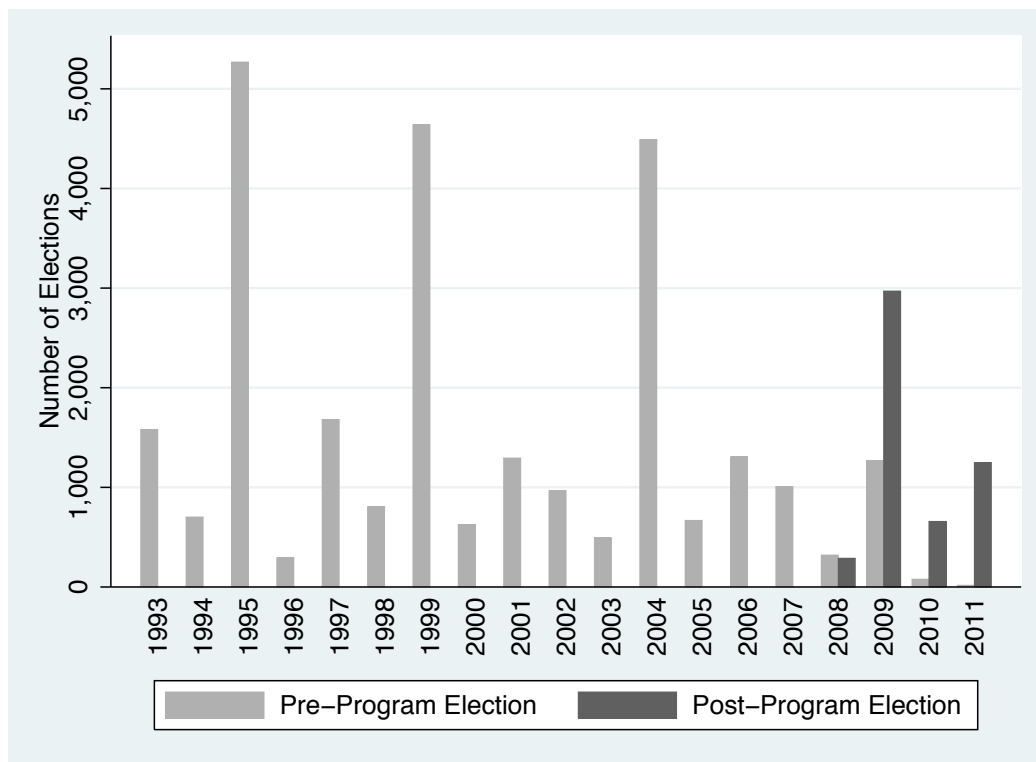
Table 2.1 presents summary statistics for the variables used in the paper. Panel A presents the main variables related to the Ghost Buildings program. Panels B and C include town-level geographical and socio-economic covariates, respectively. These are measured prior to the inception of the Ghost Buildings program, mostly in the 2001 Population Census. Panel D (“Mayor” variables”) summarizes characteristics of the mayor in office at the time of the program inception in the town.³⁰ for each Italian city in our sample. In Panel E, we summarize the local election panel variables.³¹ Tables B.1 and B.2 provide a detailed

²⁹The Italian municipal government (*Comune*) is composed of a mayor (*Sindaco*), an executive committee (*Giunta*) appointed by the mayor, and an elected city council (*Consiglio Comunale*).

³⁰Only about a half of mayors are matched to national parties. We therefore choose not to focus on this variable in our analysis. Including dummies for political alignment among the controls in the regressions does not affect the results we present later in the paper.

³¹Given that our main outcome of interest is the probability of reelection, Table 2.1 summarizes the variables

Figure 2.2: *Number of Elections per Year*



Notes: The figure shows, for each calendar year, the number of elections held before and after the inception of the Ghost Buildings program. The figure includes elections from 1993 to 2011.

description of data sources and variable definitions.

2.4.2 The Correlates of Tax Evasion

We use data from the Ghost Buildings program to study the correlates of tax evasion at the town level. Figure 2.3 presents our measure of ghost building intensity across Italian towns.³² Tax evasion is more prevalent in Southern Italy, and it is less widespread in the North. Table 2.2 presents the correlates of ghost building intensity (per 1,000 of land registry parcels). In Column (1), we first study whether geographical factors (altitude, area size of the municipality, number of land registry parcels) are correlated with tax evasion. In Column (2), we add socio-economic controls (population, income per-capita, social capital, number of firms, urbanization rate). Finally, in Column (3), we show that our results are unaffected by the inclusion of regional fixed effects.³³

We find that several geographic characteristics are strongly associated with tax evasion. In particular, controlling for other variables, tax evasion is higher in more widespread municipalities. Plausibly, in cities with wide geographical extension, the opportunities for unregistered buildings are higher as the enforcement of building registration is more difficult and resource-intensive. However, we cannot decisively interpret this evidence as causal. Previous literature has shown for example that borders are endogenously determined (see, among others, Alesina and Spolaore (1997); Alesina, Baqir and Hoxby (2004); Alesina, Easterly, and Matuszeski (2011)).

Finally, as expected, tax evasion is negatively associated with both social capital and income. In particular, the finding on social capital is consistent with Putnam (2001), who finds that the percentage of tax evasion, as measured by the Internal Revenue Service,

only for the elections in which the mayor does not face a binding a term limit.

³²For presentation purposes, we choose to show the results of this section using a measure of ghost building intensity obtained by normalizing the number of parcels with ghost buildings (per 1,000 land registry parcels).

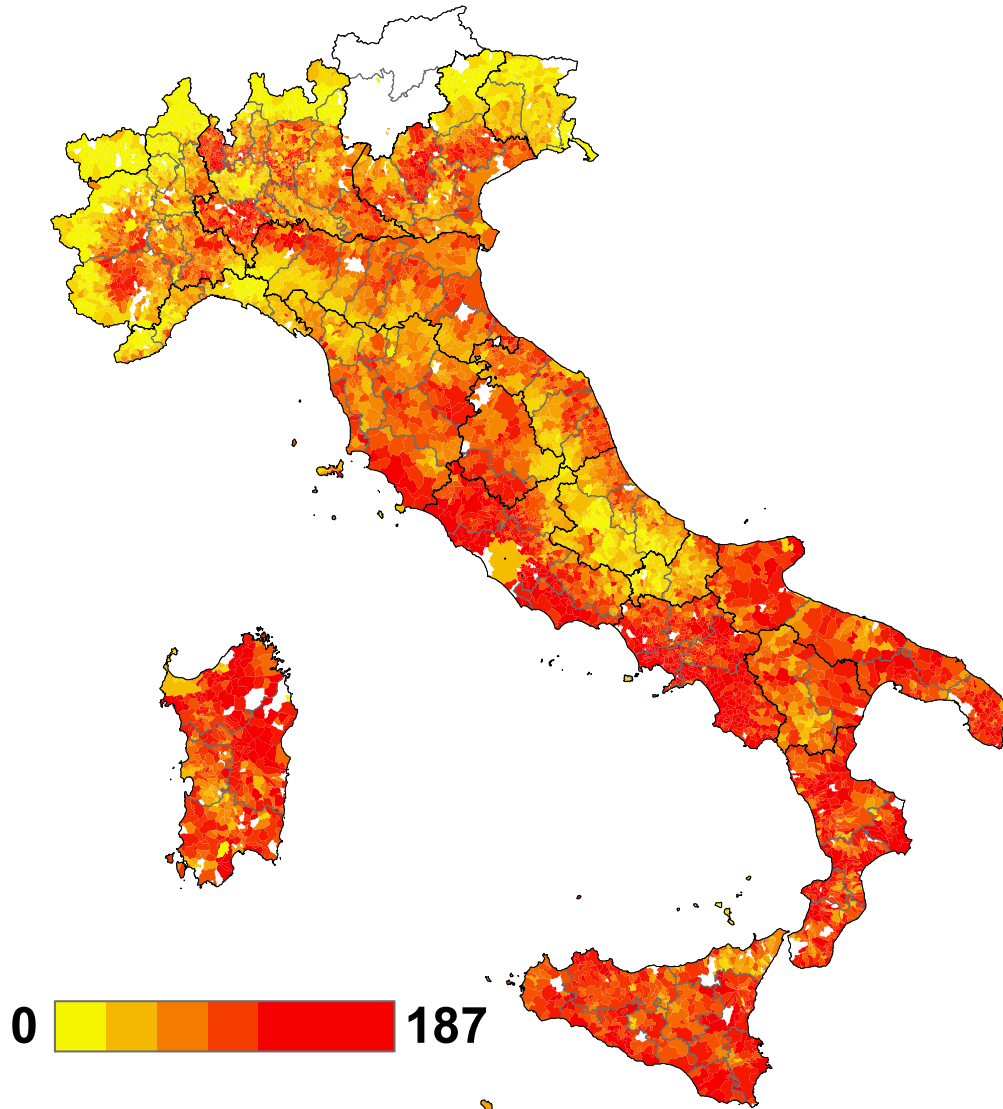
³³For 3.5% of the towns in our sample we are missing at least one town-level control. In our regressions throughout the paper, for each control, we include a binary indicator which is equal to one if the control is missing. In addition, we replace missing values with an arbitrary unique value. The results of our paper are unchanged when we undo this and just drop observations with missing values for the control variables.

Table 2.1: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Panel A: Ghost Building Town Variables					
Ghost Building Intensity	0.027	0.021	0	0.187	7720
Registered Ghost Building Intensity (Apr 2011)	0.006	0.006	0	0.051	7720
Ghost Building Registration Rate (Apr 2011)	0.243	0.181	0	1	7720
Panel B: Geographic Town Variables					
Town Area Size (sq km)	37.045	50.112	0.2	1307.71	7714
Altitude (mt)	510.535	461.655	0	3072.5	7714
Land Registry Parcels (1,000)	10.776	13.278	0.001	514.372	7720
Panel C: Socio-Economic Town Variables					
Population (1,000)	7.227	40.248	0.033	2546.804	7713
Disposable Income per capita (1,000 Euros)	13.487	3.05	5.013	44.949	7595
Urbanization Index	1.619	0.684	1	3	7713
Non-Profit Associations/1,000 pop	5.285	3.972	0.212	81.218	7461
Number of Firms per capita	0.077	0.027	0.018	0.344	7595
Panel D: Mayor Variables					
Mayor Age	49.03	9.68	21	83	7427
Mayor Education	5.36	0.78	3	9	7257
Mayor Born Same City (0/1)	0.47	0.5	0	1	7421
Mayor Term Number	1.3	0.46	1	2	7608
Mayor Woman (0/1)	0.1	0.3	0	1	7603
Panel E: Election Panel Variables					
Term Limit Indicator (0/1)	0.202	0.401	0	1	32501
Post Program Election (0/1)	0.145	0.352	0	1	25952
Years Elapsed since Program Inception (= 0 if ≤ 0)	0.306	0.816	0	4	25952
Incumbent Reelection (0/1)	0.453	0.498	0	1	25952
Incumbent Rerun (0/1)	0.572	0.495	0	1	25580
N: Candidates	2.763	1.303	1	17	24637
Victory Margin	25.953	26.938	0	100	23985
Runoff (0/1)	0.526	0.499	0	1	2297

Notes: **Socio-Economic Town Variables** are collected before the Ghost Buildings program inception. **Mayor Variables** refer to characteristics of the incumbent mayor at the time of program inception. Summary statistics for the **Election Panel Variables** are reported for the subsample of elections with no binding term limit, except for *Term Limit Indicator*. A detailed description and source of each variable is provided in the *Data Appendix*

Figure 2.3: *Ghost Building Intensity (per 1,000 land registry parcels)*



Notes: In this figure, Ghost Building Intensity is defined as the number of land registry parcels with ghost buildings per thousand of land registry parcels. White areas identify towns for which we do not have data.

Table 2.2: *The Determinants of Ghost Building Intensity (per 1,000 land parcels)*

	(1)	(2)	(3)
Town Area Size (sq km)	0.102*** (0.021)	0.123*** (0.015)	0.098*** (0.012)
Altitude (mt)	-0.015*** (0.002)	-0.009*** (0.002)	-0.011*** (0.002)
Land Registry Parcels (1,000)	-0.236*** (0.065)	-0.328*** (0.070)	-0.270*** (0.047)
Population (1,000)		0.020 (0.016)	0.004 (0.011)
Disposable Income per capita (1,000 Euros)		-2.601*** (0.363)	-1.223*** (0.291)
Urbanization Index		5.948*** (1.839)	4.407*** (1.652)
Non-Profit Associations/1,000 pop		-0.458*** (0.144)	-0.211*** (0.074)
Number of Firms per capita		56.100*** (20.040)	89.945*** (17.842)
Region FE			X
Observations	7720	7720	7720

Notes: The dependent variable is the town-level ghost building intensity per thousand of parcels, defined as the ratio between the number of land registry parcels with ghost buildings and the total number of land registry parcels, multiplied by one thousand. Standard errors are clustered at provincial level. *p<0.1, **p<0.05, ***p<0.01.

is strongly related to differences in social capital at the state level. Additionally, when analyzing the DDB Chicago Lifestyles survey, he argues that by far the best predictor of tax evasion is the number of times in the course of the last year that respondents gave the “finger” to another driver.

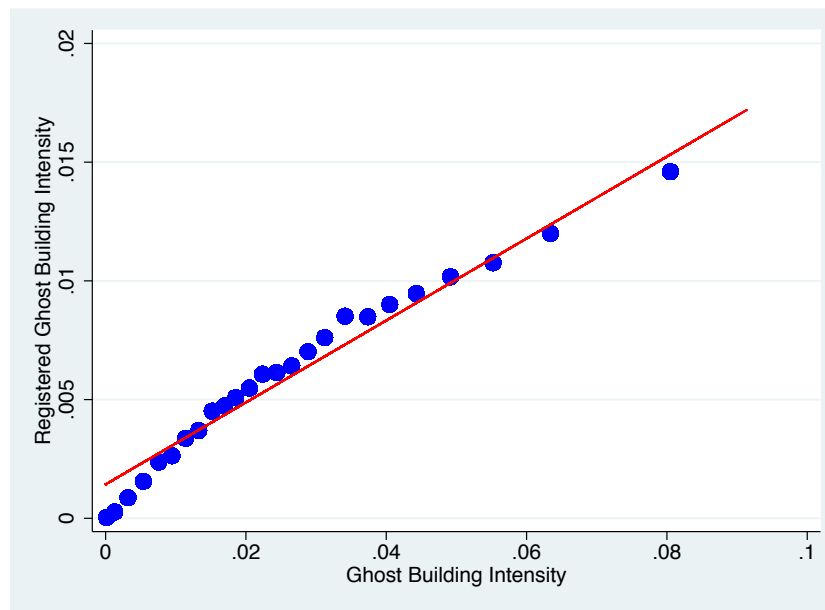
2.4.3 The Political Determinants of the Ghost Buildings Registration

We now provide more details on the wave of registration of ghost buildings induced by the program. First, we show that the number of ghost buildings detected by the program is a good predictor of the number of ghost buildings that were registered in response to the policy. Figure 2.4 displays the relation between the number of land parcels with ghost buildings eventually registered by the April 2011 deadline (“registered ghost building intensity”) and the number of parcels that were identified as containing ghost buildings (“ghost building intensity”), both as a share of the total number of land registry parcels. In the graph, the x-axis variable is partitioned into 25 quantiles. The scatter plot shows a clear increasing relation. In a linear regression analysis, an increase of one standard deviation in the detected intensity of ghost buildings raises the *registered* ghost building intensity at April 2011 by approximately 0.75 standard deviations ($p < 0.01$).³⁴ To summarize, the program scope at the town-level strongly predicts the actual impact of the program on tax enforcement. This premise motivates the strategy that we introduce in Section 2.5 to estimate the impact of the Ghost Buildings program on electoral outcomes.

Second, we analyze the ghost buildings registration rate, defined as the percentage of ghost building parcels that get registered by the April 2011 deadline. Figure 2.5 summarizes the ghost building registration rate and documents a substantial dispersion across towns. Table 2.3 documents the impact of characteristics of the mayor at the time of the program inception on this outcome. For a given level of the other covariates, the registration rate is higher when mayors are male, younger, more educated, or are born in the same city in which

³⁴The relation is basically unchanged when adding town-level controls and regional fixed effects (results available on request).

Figure 2.4: *Registered Ghost Building Intensity*



Notes: The scatter plots the relation between *Registered Ghost Building Intensity* (i.e., the fraction of land parcels with ghost buildings that get registered by April 2011) and *Ghost Building Intensity* (i.e., the fraction of land parcels with ghost buildings identified by the program). The x-axis is partitioned into 25 quantiles. The x-axis of each dot is the median value of the ghost building intensity in the quantile. The y-axis is the average value of the registered ghost building intensity in the quantile. We cut the top 1% of the x-axis values from the graph. The line plots the predicted values from a linear regression model.

they serve as mayor. The correlation between gender and policies in Italian municipality is potentially consistent with the results of Gagliarducci and Paserman (2012), who find that female policymakers usually face more difficulty in implementing policies when in office. To the extent education can be considered a proxy for politicians' quality (see, for example, Besley, Montalvo and Reynal-Querol (2011)), this set of results also supports the view that better policymakers fight tax evasion more. We highlight the correlation between the mayor's place of birth and the tax evasion enforcement. One possible explanation could be that mayors who are born in the same city have access to additional information that can facilitate tax evasion enforcement. We acknowledge that this evidence relies on cross-sectional correlation analysis and thus should be interpreted with caution. However, we also notice that the results are robust to the inclusion of geographical and socio-economic controls, in Columns (2) and (3), respectively. With these caveats in mind, the findings of this section suggest that mayor's characteristics did have a role in shaping registration activities across towns.

2.5 Empirical Strategy

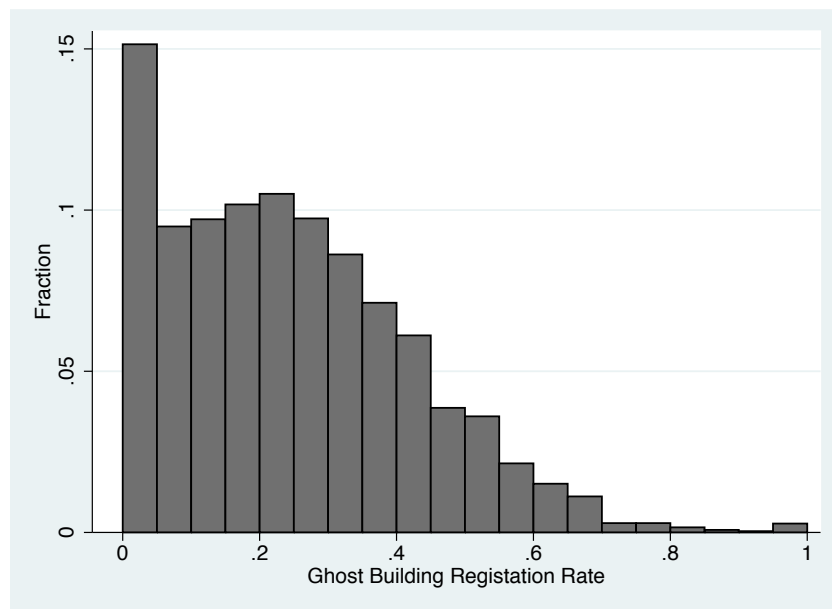
2.5.1 The Electoral Response to the Ghost Buildings Program

In this section, we outline our approach to estimate the voter response to the Ghost Buildings program. We also aim to isolate the channels that drive this response. Our empirical strategy exploits variation across towns in the program scope to increase tax enforcement.³⁵ We implement a difference-in-differences approach based on the town-level ghost building intensity, which we defined above as the ratio between the number of land registry parcels with ghost buildings and total number of land registry parcels in each town.

In Section 2.4.3, we documented that mayor characteristics, such as age, education, and

³⁵Importantly, the *Agenzia del Territorio* conducted the detection activities homogeneously throughout the country. Thus, heterogeneity in the number of detected unregistered buildings captures differences in of actual levels of non-registration at the time of the aerial photographing as opposed to differential intensity in the detection activity.

Figure 2.5: *Ghost Building Registration Rate*



Notes: The histogram shows the distribution of the ghost building rate at April 30, 2011, defined as the ratio between the number of land registry parcels ghost buildings that get registered by April 2011 and the number of land registry parcels with ghost buildings identified by the program.

Table 2.3: *The Determinants of the Ghost Building Registration Rate*

	(1)	(2)	(3)
Mayor Age	-0.067*** (0.022)	-0.068*** (0.021)	-0.057*** (0.020)
Mayor Education	0.704*** (0.265)	0.630** (0.262)	0.702*** (0.249)
Mayor Born Same City (0/1)	1.051** (0.424)	1.145*** (0.423)	0.968** (0.408)
Mayor Term Number	-0.174 (0.362)	-0.047 (0.353)	-0.041 (0.364)
Mayor Woman (0/1)	-0.923 (0.642)	-1.246** (0.626)	-1.202* (0.607)
Geographic Controls		X	X
Socio-Economic Controls			X
Observations	7720	7720	7720

Notes: The dependent variable is the town-level ghost building registration rate, defined as the ratio between the number of land registry parcels with ghost buildings that get registered by April 2011 and the number of land registry parcels with ghost buildings identified at the beginning of the program. Refer to Table 2.1 for a description of the *Geographic* and *Socio-Economic* Controls. All the regressions include regional fixed effects and year-of-program-inception fixed effects. Standard errors are clustered at provincial level. *p<0.1, **p<0.05, ***p<0.01.

gender, predict the registration rate of the detected ghost buildings. In addition, the actual levels of registration might depend on voter preferences and responsiveness to the program. Thus, a naive analysis looking at the relationship between actual ghost building registrations and reelection outcomes will suffer from the standard omitted variable bias. This motivates our focus on *ex ante* program scope to increase enforcement.

The rationale for our identification approach is that program scope at the town level predicts the increase in enforcement induced by the Ghost Buildings program, as shown in Figure 2.4. Towns with a higher share of parcels with *detected* ghost buildings also have, on average, a higher share of parcels with *registered* ghost buildings, as measured in April 2011.³⁶

³⁶Importantly, we argue that the intensity of ghost buildings is not a valid instrument for actual registration intensity. In principle, as we discussed in Section 2.1, the program could affect incumbent reelection probability through other channels besides registration. This would make the standard exclusion restriction required for an instrumental variable approach invalid.

Our baseline specification is therefore:

$$R_{imet} = \beta_0 + \beta_1 Post_{ie} \cdot Ghost\ Building\ Intensity_i + \eta_m \cdot Post_{ie} + \phi_i + \phi_t + \epsilon_{imet} \quad (2.7)$$

The dependent variable R_{iret} is a dummy that indicates whether the incumbent of municipality i in macro-area m is re-elected in election e in year t . Observations where the incumbent cannot be reelected because of a binding term limit are excluded from the regression sample. The dummy $Post$ is equal to one when election e occurs after the beginning of the Ghost Buildings program in the town. The coefficients η_m capture post-program period fixed effects that are specific to the four Italia macro-areas³⁷ m where town i is located. We also include town fixed effects, ϕ_i , and election year fixed effects, ϕ_t . Finally, *Ghost Building Intensity_i* is the intensity of ghost buildings in town i . The coefficient of interest, β_1 , thus captures the differential impact of the Ghost Buildings program on incumbent reelection by ghost building intensity. Throughout the paper, we cluster standard errors at the provincial level to allow for spatial correlation in the error term.

We adopt a similar regression model to study the impact of the program on other electoral competitiveness outcomes. We focus on four variables: i) a binary indicator for whether the incumbent reruns for election; ii) the number of candidates running for the mayor office; iii) the difference in the percentage of votes between the first and the second candidate;³⁸ iv) a binary indicator equal to one if a runoff takes place, which occurs in towns with more than 15,000 inhabitants when none of the candidates obtain the absolute majority in the first-round.

One potential challenge to our identification strategy may arise from the town-specific timing of publications of the unauthorized building lists. On the one hand, if local administrators had influence over publication date, unpopular mayors in cities with high evasion rates might lobby to delay the publication. On the other hand, the central government might push to start the program earlier in those towns where mayors set a lower level of

³⁷North, Center, South, Islands.

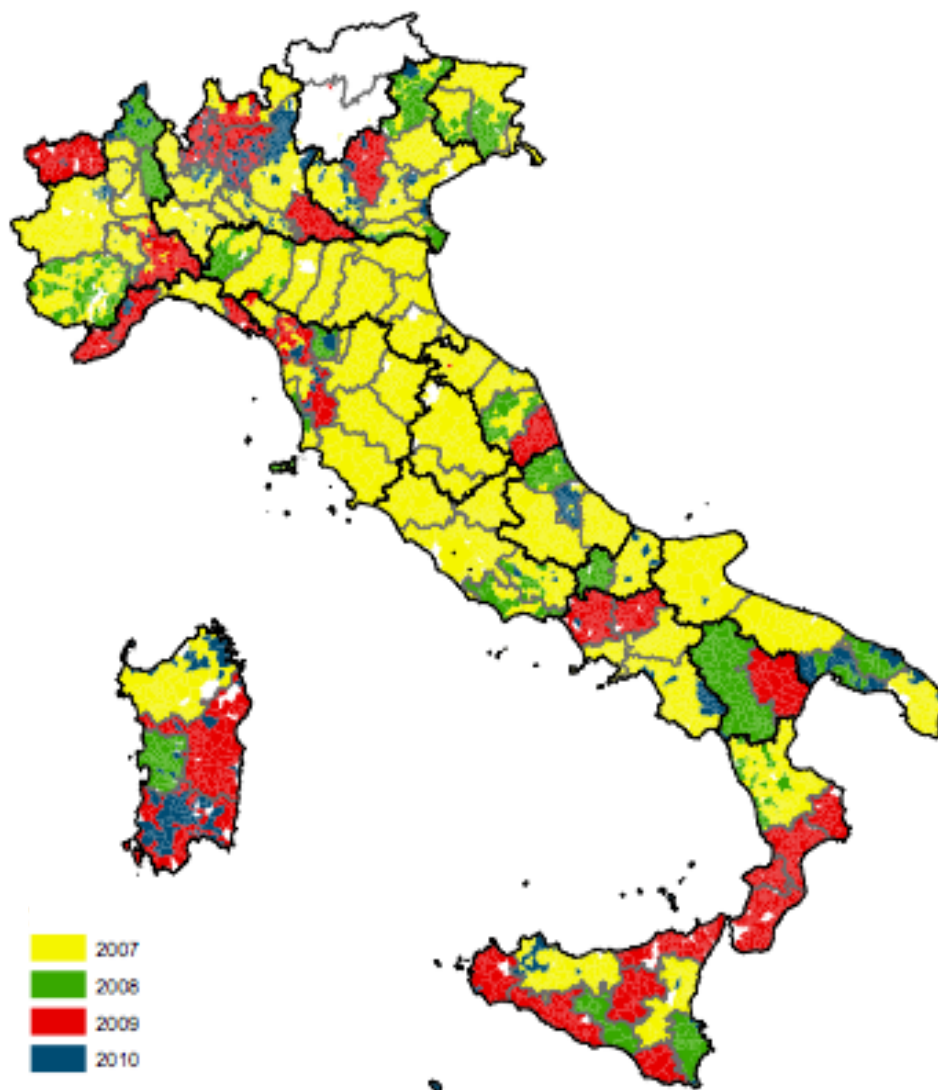
³⁸For elections with a runoff, we use the first round results.

tax enforcement. In both these cases, our estimates of the impact of the Ghost Building program on reelection likelihood might just capture a selection effect. We handle this concern in several ways. First, as discussed in Section 2.2, we notice that the timing of the publication was primarily determined at the provincial level by the availability of digital land registry maps and was highly clustered by province. Figure 2.6 emphasizes the high level of provincial clustering in the publication years. Only about 7% of the post-program elections have values for the post-program indicator different from the one they would have had based on the modal date of publication in the province. To deal with these discrepancies we implement an instrumental variable approach. We code elections based on whether they occur before or after the modal date of publication of the unauthorized building lists in the province. We then instrument the actual *Post* dummy with this binary indicator at the provincial level. We adopt this strategy for our main specifications.³⁹ In addition, in Section 2.6.2, we present robustness checks using an alternative instrument for the post-program indicator using the *national* modal program inception year.

As is standard in difference-in-differences estimation, identification of the coefficient of interest relies on two assumptions. The first is the absence of contemporary events that differentially affected towns having higher ghost building intensity. We are not aware of other policies targeting this form of tax evasion happening concurrently with the Ghost Buildings program. However, it is still possible that other events, which differ in intensity by other variables correlated with ghost building intensity, occurred at the same time. We address the concern by presenting alternative specifications where we include interactions between a comprehensive set of geographical, socio-economic, and political controls, all measured before the beginning of the program, and the post-program binary indicator. The second assumption is the presence of parallel trends in the outcome variable. We assess this assumption using several tests and placebo exercises in the next section.

³⁹The towns targeted by the program belonged to 101 provinces.

Figure 2.6: *Ghost Buildings Program Inception Year*



Notes: The figure shows the year of inception of the Ghost Building program (i.e., the year of publication of the list of ghost buildings) in each town. White areas identify towns for which we do not have data.

2.5.2 Channels

The reduced form approach presented thus far tests whether higher program scope to increase tax enforcement at the town level affects incumbent reelection likelihood in the post-program period. We complement this baseline regression with further analysis. First, we show that it is the registration induced by the program that drives the electoral response, as opposed to other potential interpretations. For this purpose, we use actual ghost building registration data. In Section 2.4, we emphasized several important measurement limitations of these data that warrant caution. With this caveat in mind, we test whether, for a given intensity of ghost buildings, a higher ghost building registration rate (*Registration Rate*) induced by the program has a positive effect on incumbent reelection likelihood:

$$\begin{aligned} R_{imet} = & \gamma_0 + \gamma_1 Post_{ie} \cdot Ghost\ Building\ Intensity_i \\ & + \gamma_2 Post_{ie} \cdot Registration\ Rate_i + \zeta_m \cdot Post_{ie} + \mu_i + \mu_t + v_{imet} \end{aligned} \quad (2.8)$$

As we discussed above, an obvious threat to identification of γ_2 in Equation 2.8 arises from the fact that the registration effort is potentially correlated with many potential town-level confounders. We first check robustness of the results to the inclusion of mayor controls. In addition, the timing of the program provides a strategy that can alleviate this concern. Even if the program started in the same year in most of the towns, we can exploit the variation generated by the fact that Italian municipalities hold elections in different years. A longer time between the beginning of the program and the election date naturally leads to more registration activities. This generates variation across towns in the registration rate achieved prior to the local election date that is plausibly uncorrelated with mayor quality. We use this instrumental variable strategy to look at the impact of a change in registration rate on incumbent reelection likelihood.

Second, we shed light on the channels through which the program could affect voters' political preferences. Consistently with the theoretical framework, we investigate the interaction among the political returns to an increase in tax enforcement, local government efficiency in delivering public goods, and "tax culture", the stigma associated to evading

taxes. We use the speed of public good provision as a proxy for the quality of the delivery at the *municipal* level. This indicator is measured as the ratio of paid outlays in the municipal financial report over the total outlays committed in the budget. The intuition is that the provision of public goods is more effective in places where the actual allocation delivered to citizens is closer to the amount allocated in the budget.⁴⁰ The following regression model tests whether the electoral response to the Ghost Building program varies by speed of public good provision:

$$R_{iret} = \delta_0 + \delta_1 Post_{ie} \cdot Ghost\ Building\ Intensity_i + \delta_2 Post_{ie} \cdot Speed\ Public\ Good\ Provision_i + \delta_3 \cdot Post_{ie} * GB_i \cdot Speed\ Public\ Good\ Provision_i + \zeta_r \cdot Post_{ie} + \lambda_i + \lambda_t + v_{iret} \quad (2.9)$$

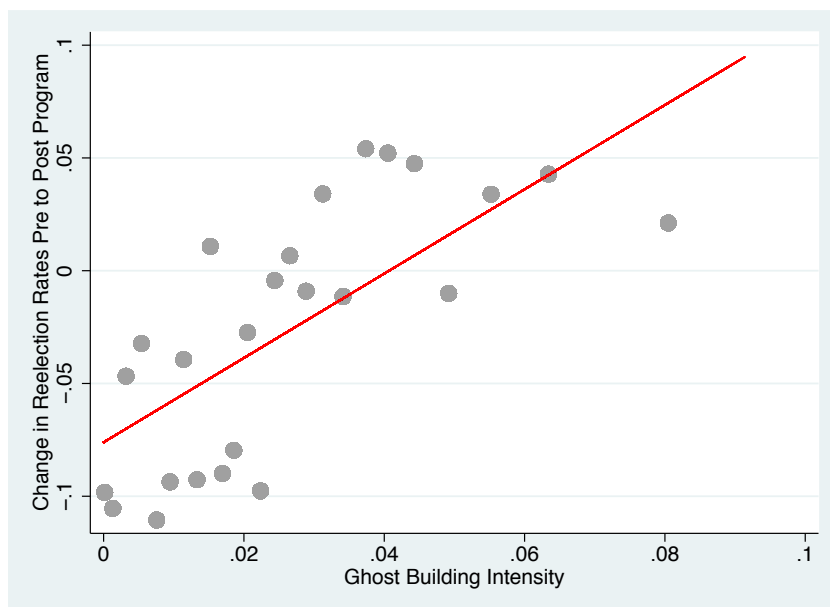
where δ_3 is the coefficient of interest.

We then use data from the *European Value Study* — the European component of *World Values Survey*— to study the role of tax culture. Specifically, we use the question: “Do you justify cheating on tax?”. For these data, geographical identification of respondents is available only at the regional level. We thus compute and standardize region-level means. The regression model to capture heterogeneity by this variable is similar to the one presented in Equation 2.9.

Finally, we also assess the impact of the program on town-level public expenditures. To test whether the program scope to increase tax enforcement affected these expenditures, we adopt a specification similar to the one we presented in Equation 2.7, using the natural logarithm of the local government expenditures as the dependent variable.

⁴⁰ A similar proxy has been used by Grembi, Nannicini and Troiano (2013) and Gagliarducci and Nannicini (2013) to measure the quality of public good delivery. For our sample, we compute the speed of public good provision as the average across two pre-treatment years. The results are similar with alternative definitions. Results are available upon request.

Figure 2.7: *Difference in reelection rates pre- to post- Ghost Buildings program*



Notes: The scatter plots the relation between the change in the average (year-demeaned) reelection rate between the pre-program and the post-program periods and the *Ghost Building Intensity* (i.e., the ratio between the number of land registry parcels with ghost buildings and the total number of land registry parcels). The x-axis is partitioned into 25 quantiles. The x-axis of each dot is the median value of the ghost building intensity in the quantile. The y-axis is the average value of the registered ghost building intensity in the quantile. We cut the top 1% of the x-axis values from the graph. The sample includes elections with no binding term limit. The line plots the predicted values from a linear regression model.

2.6 Results

2.6.1 Baseline Results

In this section, we investigate the electoral consequences of the Ghost Buildings program. Figure 2.7 provides a visual analysis of the relation between ghost building intensity and changes in the incumbent reelection likelihood — our main outcome variable — after the beginning of the program. On the x-axis, the ghost building intensity is partitioned into 25 quantiles. The scatter displays a clear increasing relation, with the quantile dots fairly closed aligned along the fitting line.

Table 2.4 formalizes the analysis above and presents the results of the difference-in-differences estimation discussed in Section 2.5. Column (1) reports the basic OLS spec-

ification (“Reduced Form”) using the provincial post-program indicator. The coefficient remains stable with the inclusion of fixed effect (Column (2)) and of election year fixed effect (Column (3)). Starting in Column (4), we instrument the post-program indicator with the provincial post-program indicator. The coefficient is stable across the different specifications. The inclusion of year fixed effects and town fixed effects, in Columns (5) and (6) respectively, does not change our results. In Column (6) — the baseline specification for the rest of the analysis — the reported coefficient on the interaction between the ghost building intensity and the post-program indicator is 1.042, significant at 1%. This magnitude implies that a one standard deviation in the town-level program scope to increase enforcement, as measured by the ghost building intensity, raises the likelihood of reelection of the incumbent by approximately 2.5 percentage points in post-program elections, relative to pre-program ones (from a sample mean of 45.4). A back of the envelope calculation suggests that the effect of a one standard deviation increase in Ghost Buildings program scope on incumbent reelection probability is on the order of magnitude of i) 6% of the incumbency effect in U.S. House Elections (Lee (2008)) and ii) the effect of a 5% increase in town government spending in Brazil municipal elections (Litschig and Morrison (2012)).

We adopt an analogous regression strategy to study the impact of the program on other measures of election competitiveness as described in Section 2.5. For each of these variables, we report the specification used in Column (6) of Table 2.4. Table 2.5 presents a clear picture. An increase in ghost building intensity raises the likelihood that the incumbent re-runs, and decreases competitiveness of the elections. Specifically, a one standard deviation increase in ghost building intensity reduces the number of candidates by 0.03 standard deviations, increases the margin of victory by 0.05 standard deviations, though this last result is not statistically significant at conventional levels, and reduces the likelihood of a runoff by 15% of the mean value. This is consistent with the idea that potential entrants in the electoral competition correctly anticipate stronger incumbent advantage in response to the program.

Table 2.4: Ghost Building Intensity and Incumbent Reelection: Baseline Results

	Reduced Form			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Ghost Building Intensity*Post				1.099*** (0.357)	1.114*** (0.373)	1.042*** (0.378)
Ghost Building Intensity*Province Post	1.085*** (0.343)	1.061*** (0.358)	0.953*** (0.360)			
Town FE		X	X		X	X
Election Year FE			X			X
Observations	25893	25893	25893	25893	25893	25893

Notes: The dependent variable is a binary indicator equal to one if the incumbent mayor is reelected (mean 0.453). **Post** is a binary indicator equal to one if the election occurs after the Ghost Buildings program inception. **Province Post** is a binary indicator equal to one if the election occurs after the Ghost Buildings program modal inception year in the province. In the columns grouped under the header "2SLS", *Post* is instrumented with *Post Province*. **Ghost Building Intensity** is defined as the ratio between the number of land registry parcels with ghost buildings and the total number of land registry parcels. All the columns include an interaction between region fixed effects and either *Province Post* (Columns (1)-(3)) or *Post* (Columns (4)-(6)). Columns (1) and (4) include the Ghost Building Intensity level. The regression sample includes all the elections between 1993 and 2011 in which the incumbent does not face a binding term-limit. Standard errors are clustered at provincial level. *p<0.1, **p<0.05, ***p<0.01.

Table 2.5: Ghost Building Intensity and Election Competitiveness

	Incumbent Rerun	N. Candidates	Victory Margin	Runoff
	(1)	(2)	(3)	(4)
Ghost Building Intensity*Post Program	1.115** (0.457)	-2.383** (1.057)	42.063 (25.984)	-4.502*** (1.383)
Dependent Variable Mean	0.572	2.761	25.999	0.525
Observations	25483	24441	23562	2216

Notes: **Incumbent Rerun** is a binary indicator equal to one when the current incumbent runs for reelection. **N. Candidates** is the number of candidates running for election. **Victory Margin** is the percentage point difference between the first and the second candidate in the elections (we use first-round percentages even for towns with a runoff). **Runoff** is a binary indicator, defined only for towns with more than 15,000 inhabitants, equal to one if the election requires a runoff. This occurs if the first candidate in the first round receives less than 50% of the votes. **Post** is a binary indicator equal to one if the election occurs after the Ghost Buildings program inception. In all the columns, *Post* is instrumented by *Province Post*, a binary indicator equal to one if the election occurs after the modal program inception year in the province. **Ghost Building Intensity** is defined as the ratio between the number of land registry parcels with ghost buildings and the total number of land registry parcels. All the regressions include town fixed effects, election-year fixed effects and an interaction between macro-areas fixed effects and *Post*. The regression sample includes all the elections between 1993 and 2011 in which the incumbent does not face a binding term-limit. The ghost building intensity is defined as the ratio between the number of land parcels with ghost buildings (as measured at the beginning of the program) and the total number of land parcels in the town. Standard errors are clustered at provincial level. *p<0.1, **p<0.05, ***p<0.01.

2.6.2 Identification Validity and Robustness Checks

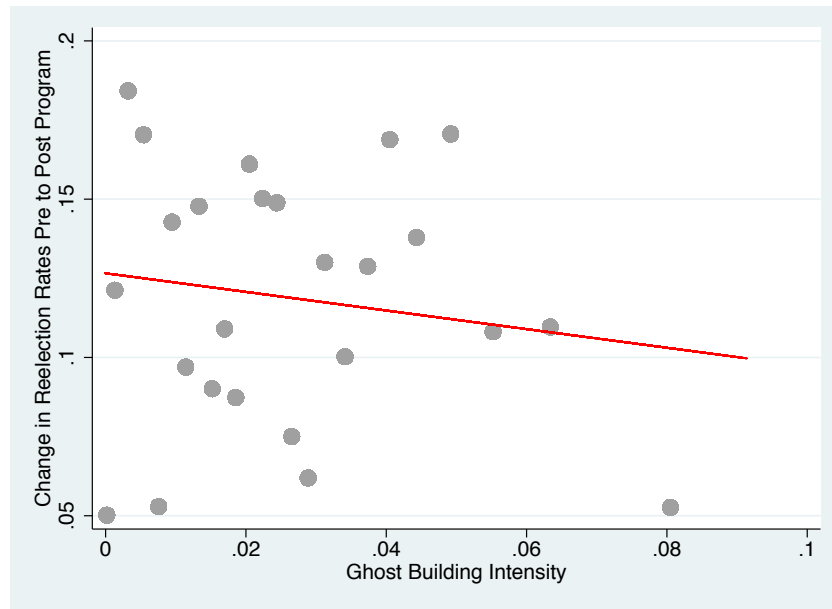
In this section, we report several auxiliary results that support the validity of our identification strategy. We focus on our main outcome variable: the likelihood of incumbent reelection.⁴¹ Figure 2.8 presents a placebo version of Figure 2.7. We look only at pre-program elections and define a placebo program year in the median year of this restricted sample. In this exercise, the relation between changes in incumbent reelection likelihood from the pre- to the post-program periods appears and ghost building intensity is very noisy and, if anything, negative.

Second, we test whether the inclusion of town-level controls affects our results. Specifically, we look at three sets of town-level variables that we present in Table 2.1: geographic features, socio-economic variables, and incumbent mayor characteristics at the time of the program commencement. We include the interaction between these variables and the post-program indicator in our regression model (the level is absorbed by the town fixed effect). If the findings in Table 2.4 were reflecting differential changes in the outcome along variables correlated to the ghost building intensity, we would expect the coefficient on the interaction between ghost building intensity and post-program indicator to be substantially affected by the inclusion of these extra controls. Table 2.6 presents the results of this analysis. We find that the baseline results are very robust to the inclusion of each set of controls both separately (Columns (2)-(4)) and together (Column (5)).

In Figure 2.9, we check whether towns with different levels of evasion were on different trends before the treatment. We report point estimates and confidence intervals on ghost building intensity for each of the elections pre- and post-program. The figure shows that, before the Ghost Buildings program started, the probability of reelection of the incumbent was independent of tax evasion. None of the pre-program coefficients are either significantly different from zero (the normalized value of the coefficient in the „Äú-1,Äù election, the omitted group) or significantly different from each other. However, after the beginning of the program there is a statistically and economically significant impact. Thus, the coefficient

⁴¹We obtain similar results for our auxiliary outcome variables. Results are available on request.

Figure 2.8: *Difference in reelection rates pre- to post- Placebo Ghost Buildings program*



Notes: The scatter plots the relation between the change in the average (year-demeaned) reelection rate between the pre-program and the post-program periods and the *Ghost Building Intensity* (i.e., the ratio between the number of land registry parcels with ghost buildings and the total number of land registry parcels). The x-axis is partitioned into 25 quantiles. The x-axis of each dot is the median value of the ghost building intensity in the quantile. The y-axis is the average value of the registered ghost building intensity in the quantile. We cut the top 1% of the x-axis values from the graph. The sample includes elections with no binding term limit. The line plots the predicted values from a linear regression model. The **placebo subsample** of observations used for this graph only includes election that occurred before the actual program inception. In each town, the year of the placebo program start is defined as nine years before the actual publication. This roughly divides the graph sample in two equally sized groups of pre-placebo and post-placebo elections. The domain of the ghost building intensity is partitioned in 25 quantiles. The x-axis of each dot is the median value of the ghost building intensity in the quantile. The y-axis is the average value of the registered ghost building intensity in the quantile. We trim the top 1% of the x-axis values. The sample includes elections with no binding term limit. The line plots the predicted values from a linear regression model.

Table 2.6: *Ghost Building Intensity and Incumbent Reelection: Additional Controls*

	(1)	(2)	(3)	(4)	(5)
Ghost Building Intensity*Post	1.042*** (0.378)	0.850** (0.370)	1.053** (0.410)	1.254*** (0.360)	1.096*** (0.374)
Geographical Controls*Post		X			X
Socio-Economic Controls*Post			X		X
Mayor Controls*Post				X	X
Observations	25893	25893	25893	25893	25893

Notes: The dependent variable is a binary indicator equal to one if the incumbent mayor is reelected (mean 0.453). **Post** is a binary indicator equal to one if the election occurs after the Ghost Buildings program inception. In all the columns, *Post* is instrumented by *Province Post*, a binary indicator equal to one if the election occurs after the modal program inception year in the province. **Ghost Building Intensity** is defined as the ratio between the number of land registry parcels with ghost buildings and the total number of land registry parcels. Refer to Table 2.1 for a description of the *Geographic*, *Socio-Economic*, and *Mayor* Controls. All the regressions include town fixed effects, election-year fixed effects and an interaction between macro-areas fixed effects and *Post*. The regression sample includes all the elections between 1993 and 2011 in which the incumbent does not face a binding term-limit. Standard errors are clustered at provincial level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

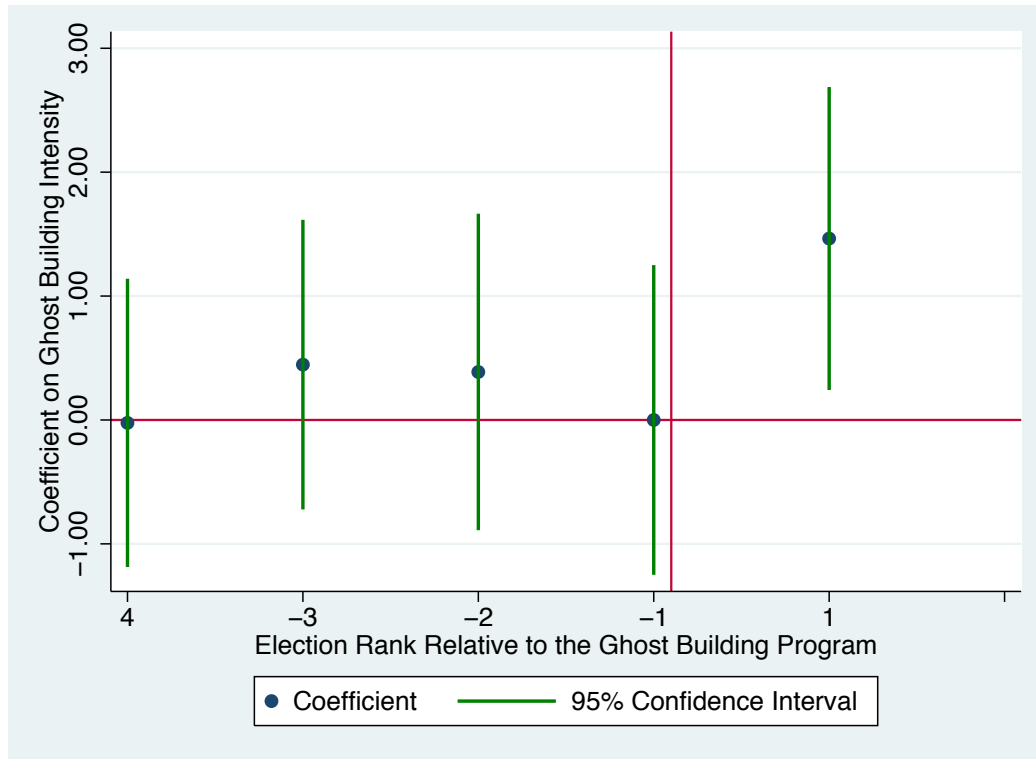
pattern in Figure 2.9 suggests that the common trend assumption holds in our setting.⁴²

Table 2.7 presents several additional robustness checks. Column (1) shows our baseline specification. In Column (2), we show that the results are robust to the inclusion between regional dummies and post-program indicators. In Column (3), we show that our results hold if we trim the ghost building intensity at the top 1 percent. If anything, the coefficient size grows. This suggests that outliers (i.e., cities with an abnormally large fraction of unregistered buildings) are not driving the results. In Column (4), we check robustness of the results to dropping small towns with a low number of land registry parcels (we drop the bottom 10%). One might be concerned that our results are driven by cities with a large number of non-residents who own a building. If this were the case, mayors would not pay the electoral cost of enforcing tax evasion. Therefore, we exclude in Column (5) cities that are classified as tourist destinations by *Ancitel*, the Italian Association of Cities.⁴³ The

⁴²Appendix Figure B.1 shows that the parallel trend assumptions hold also for the other political outcome variables described above.

⁴³Ancitel defines as tourist destinations those cities with a large percentage of touristic income over the total city income. As noted by the Italian Association of Real Estate Agents (*FIAIP*) in their yearly report cities with touristic activities are usually those with a large number of buildings not owned by residents.

Figure 2.9: *Ghost Building Intensity Coefficient by Election Pre/Post Program*



Notes: The graph reports the coefficients on the ghost building intensity for each election before and after the beginning of the Ghost Buildings program. On the x-axis, elections are ranked based on their occurrence relative to the program. The regression includes town and year fixed effects. For each election rank, we report the point estimate and the 95% confidence interval. The last election before the program ("-1") is the omitted category. The coefficient on ghost building intensity for this election is normalized to zero. Confidence interval width for this election is obtained as the mean of the confidence interval width in election -2 and election +1. The modal number of years between elections is five years between 1993 and 2001, and four afterwards.

results are robust to this sample restriction. Column (6) reports an alternative definition of the post-program indicator. We instrument the town-level indicator with a binary indicator that takes value one for the years 2007-2011. This approach treats 2007, the first and modal program start year in the country, as the *intended* program start year for all the towns. It thus estimates the Local Average Treatment Effect for those towns that started the program in that year, which constitutes a different population of compliers relative to the main specification. The coefficient is about one standard-error larger than our baseline specification and still significant at 1%. Finally, we report an alternative normalization for our dependent variable. As the data reported by the *Agenzia del Territorio* included the number of parcels with unregistered buildings, we divided this outcome by the total number of land registry parcels in each municipality in our baseline specification. In order to show that this choice does not affect the result, in Column (7) we estimate our equation using the total number of buildings recorded in the town as a normalizing factor, rather than the total number of land registry parcels. We get very similar results. The effect of one standard deviation in this alternative variable is comparable in magnitude to the one obtained when using our main ghost building intensity variable.⁴⁴

2.6.3 Channels

This section elaborates on some of the potential channels through which the anti tax evasion program could increase voter support for the incumbent. Table 2.8 presents the results from the estimation of Equation 2.8. This step aims to show that the increase in tax enforcement induced by the program — the ghost buildings registration — drove the electoral response.

In Columns (1) and (2), we present the correlation between the registration rate and the incumbent reelection. We find that, controlling for ghost building intensity, a one standard deviation increase in ghost building registration rate raises reelection likelihood by 1.3 percentage points. The result holds when using both a) the April 2011 registration rate with year fixed effects and b) the registration rate reached by the election year computed

⁴⁴Appendix Figure B.2 also shows that the results are robust to changes in the election sample years.

Table 2.7: Ghost Building Intensity and Incumbent Reelection: Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Baseline		Region Interactions	Trim top 1%	Drops Small Towns	Drop Touristic Towns	Post 2007	Alternative Normalization
Ghost Building Intensity*Post	1.042*** (0.378)	1.162*** (0.386)	1.255** (0.493)	1.195*** (0.390)	0.911** (0.416)	1.345** (0.563)	0.119* (0.067)
Observations	25893	25893	25618	23337	22905	25893	25893

Notes: The dependent variable is a binary indicator equal to one if the incumbent mayor is reelected (mean 0.453). **Post** is a binary indicator equal to one if the election occurs after the Ghost Buildings program inception. In all the columns, *Post* is instrumented by *Province Post*, a binary indicator equal to one if the election occurs after the modal program inception year in the province. All the regressions include town fixed effects, election-year fixed effects and an interaction between macro-areas fixed effects and *Post*. In the specification **Region Interactions**, we add interactions between the post-program indicator and a regional fixed effect. In the specification **Trim top 1% Intensity**, we drop towns in the first percentile of ghost building intensity (0.0928). In the specification **Drop Small Towns**, we exclude from the sample towns within the bottom 10% of the distribution of land registry parcels (2633). In the specification **Drop Touristic Towns**, we drop towns that are classified as tourist destinations by the Italian Association of Cities, *Ancitel*. In Columns (1)-(3) and (5) *Post* is instrumented by *Province Post*, a binary indicator equal to one if the election occurs after the modal program inception year in the province. In Column (4) — **Post 2007** — we instrument it with a binary indicator equal to one if the election occurs after 2007, the start date for the first (and modal) round of the program in the country. **Ghost Building Intensity** is defined as the ratio between the number of land registry parcels with ghost buildings and the total number of land registry parcels. except in Column (5) — **Alternative Normalization** — where it is defined as the ratio between the number of land registry parcels with ghost buildings and the total number of buildings. All the regressions include town fixed effects, election-year fixed effects and an interaction between macro-areas fixed effects and *Post*. The regression sample includes all the elections between 1993 and 2011 in which the incumbent does not face a binding term-limit. Standard errors are clustered at provincial level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

as described in Section 2.4. In Column (3), we show that adding the interaction between town- and mayor-level controls and the post-program indicator does not change the result. In Column (4), we show that the motivation for the instrumental variable strategy for the registration rate finds support in the data: years elapsed since the program start at election time are a good predictor of the registration rate at that time. In Column (5) we use the years elapsed since the program start as an instrument for registration rate.⁴⁵ In the IV specification, a one standard deviation increase in the registration rate (.079) raises the reelection likelihood by 4 percentage points in post-program elections. In Column (6) we show that the IV estimate is unchanged when adding town and mayor controls. Finally, Column (7) shows that an alternative computation of the registration rate imputable to the incumbent — assuming a constant growth rate of 50% in the registration levels — delivers similar results.

In Columns (5) and (7), we notice that the IV estimates are larger than the respective OLS estimates. This is relatively common (see, for example, the returns to schooling literature, reviewed in Card (2001)), and it can be explained either by OLS attenuation bias due to measurement error, or by the fact that in the set of cities affected by the IV — that is, cities where the registration activity depends on program duration — the political returns to registration might be bigger than in the rest of the cities (i.e., we are estimating a LATE). Even if our instrument is uncorrelated with any idiosyncratic city-specific characteristics, we are not able to rule out the possibility that having the program for longer time has an independent effect on its impact on the probability of reelection. While we acknowledge this possibility, we still believe that our instrument does a good job in addressing the main endogeneity concern for the registration efforts of the mayors (town-specific characteristics, such as the mayor's ability or effort).

We then provide empirical support for two channels affecting the electoral response

⁴⁵In our IV specification we do not control for year fixed effects. Three quarters of the post-program elections come from cities that started the program in 2007. Thus, we lose statistical significance when running this specification, though it is reassuring that the coefficient of interest remains of similar size. Results are available upon request.

Table 2.8: Ghost Building Registration and Incumbent Reelection

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	OLS	1st Stage	2SLS	2SLS	2SLS
Ghost Building Registration Rate*Post	0.173** (0.072)	0.220*** (0.082)	0.199*** (0.076)		0.625*** (0.196)	0.654*** (0.197)	0.680*** (0.214)
Years Elapsed since Program Inception				0.066*** (0.006)			
Ghost Building Intensity*Post	1.276*** (0.383)	1.271*** (0.374)	1.367*** (0.410)	-0.734** (0.312)	1.561*** (0.391)	1.877*** (0.445)	1.616*** (0.397)
Registration Rate	April2011	Constant	Constant	Constant	Constant	Constant	Growth
Extra Controls*Post			X			X	
Observations	25893	25893	25893	25893	25893	25893	25893

Notes: The dependent variable is a binary indicator equal to one if the incumbent mayor is reelected (mean 0.453), except in Column (3), First Stage, where it is the ghost building registration rate. **Registration Rate** refers to alternative definitions of the ghost-buildings registration rate we use in the analysis. The variable is equal to "April 2011" when we use the registration rate at April 2011 and introduce an election year fixed effects in the model. It is equal to "Constant" when we impute the registration rate at the time of the election using the April 2011 rate as a starting point and assuming a constant yearly registration rate. It is equal to "Growth" when we impute the registration rate at the time of the election using the April 2011 rate as a starting point and assuming yearly registration rate grows by 50% every year. The variable **Years Elapsed since Program Inception** takes value 0 if the elections occurs before the program. **Post** is a binary indicator equal to one if the election occurs after the Ghost Buildings program inception. In all the columns, *Post* is instrumented by *Province Post*, a binary indicator equal to one if the election occurs after the modal program inception year in the province. **Ghost Building Intensity** is defined as the ratio between the number of land registry parcels with ghost buildings and the total number of land registry parcels. **Extra Controls*Post** include *Geographic*, *Socio-Economic*, and *Mayor Controls* interacted with the *Post* dummy. Refer to Table 2.1 for a list of these variables. All the regressions include town fixed effects, election-year fixed effects and an interaction between macro-areas fixed effects and *Post*. The regression sample includes all the elections between 1993 and 2011 in which the incumbent does not face a binding term-limit. Standard errors are clustered at provincial level. *p<0.1, **p<0.05, ***p<0.01.

to the program. Our simple theoretical framework predicted that this should be higher in towns where the local government delivers public goods more effectively and where the non-monetary returns to tax enforcement are higher. We provide evidence about these hypotheses by estimating Equation 2.9. The coefficient of interest δ_3 captures the impact of a standard-deviation increase in the variables measuring either the municipal speed of public good provision or the tolerance for tax evasion on the electoral response to the program. Table 2.9 presents the results. In column (1) we find that a one standard deviation increase in the speed of public good provision increases the point estimate of the impact of ghost buildings on reelection by 0.63, and that this coefficient is statistically significant at the ten percent level. We then confirm that this interaction effect does not simply capture geographical variation in the responsiveness across different parts of Italy by adding triple interactions across the post-program indicator, the ghost building intensity, and the macro area dummies. The sign and economic significance of the coefficient is robust, though estimated less precisely ($p=.137$).

We then look at the role of tax culture. We exploit variations across regions in the extent to which respondents “justify tax cheating” in the *European Values Study*. These results provide clear evidence the tax culture matters. In Column (3), we show that a one standard deviation increase in the tolerance score reduces the point estimate of the impact of ghost buildings on reelection by .64, (significant at the ten percent level). Column (4) shows that the magnitude of the coefficient is stable, or if anything increases (in absolute value) when adding the triple interactions with macro-areas dummies. These results provide suggestive evidence that the positive effect of the Ghost Buildings program on incumbent reelection likelihood is larger in localities where voters have, on average, stronger preferences for tax enforcement and where the delivery of public goods is more effective.

Finally, Table 2.10 presents the results of the estimation of the baseline regression model in Equation 2.7, using the log of town-level government expenditures. Column (1) presents the reduced-form results, using the post-program indicator based on the provincial mode. The point estimate is .436 (significant at 10%). The coefficient is stable when instrumenting

Table 2.9: Ghost Building Intensity and Incumbent Reelection: Heterogeneity Analysis

	(1)	(2)	(3)	(4)
Ghost Building Intensity*Post	1.183*** (0.394)	1.232* (0.684)	1.063*** (0.380)	1.311** (0.668)
...*Speed of Public Good Provision	0.627* (0.380)	0.591 (0.398)		
...*Justify Tax Cheating			-0.639* (0.364)	-0.734* (0.404)
GBI*Macro Area*Post	No	Yes	No	Yes
Observations	25812	25812	25893	25893

Notes: The dependent variable is a binary indicator equal to one if the incumbent mayor is reelected (mean 0.453). **Post** is a binary indicator equal to one if the election occurs after the Ghost Buildings program inception. In all the columns, *Post* is instrumented by *Province Post*, a binary indicator equal to one if the election occurs after the modal program inception year in the province. **Ghost Building Intensity** is defined as the ratio between the number of land registry parcels with ghost buildings and the total number of land registry parcels. **GBI*Macro Area*Post** is the triple interaction among macro-areas fixed effect, ghost building intensity and *Post*. All the regressions include town fixed effects, election-year fixed effects, interactions between macro-areas fixed effects and *Post*, and an interaction between the relevant heterogeneity variable for the column and *Post*. The regression sample includes all the elections between 1993 and 2011 in which the incumbent does not face a binding term-limit. Standard errors are clustered at provincial level. *p<0.1, **p<0.05, ***p<0.01.

the post-program indicator with the provincial one and it is slightly larger when including interactions among controls and the post-program indicator.⁴⁶ While the effect of the program is statistically significant, we also note that it is fairly small. A one standard-deviation increase in ghost-building intensity increases expenditures by about 1%. We argue that it is unlikely that this effect explains the whole incumbent reelection effect we documented earlier in the paper. Consistently with the suggestive evidence provided by the heterogeneity by tax culture, we suggest that non-monetary factors (e.g., the direct utility non-evaders derive from catching of the shirkers) must play an important role.

2.6.4 Alternative Explanations

Finally, we use the entire set of our results to argue that the impact on incumbent reelection probability arising from the increase in tax enforcement more than offsets several alternative

⁴⁶Appendix Figure B.1 shows that government expenditures satisfy the parallel trend assumption, and pre-program coefficients are not statistically different from zero.

Table 2.10: *Local Government Expenditures*

	OLS	2SLS	
	(1)	(2)	(3)
Ghost Buildings Intensity * Post		0.488*	0.607***
		(0.264)	(0.207)
Ghost Building Intensity*Post Province	0.436*		
	(0.252)		
Extra Controls*Post	No	No	Yes
Observations	74646	74646	74646

The dependent variables is the natural logarithm of town government expenditures. *Notes:* **Post** is a binary indicator equal to one if the election occurs after the Ghost Buildings program inception. In all the columns, *Post* is instrumented by *Province Post*, a binary indicator equal to one if the election occurs after the modal program inception year in the province. **Ghost Building Intensity** is defined as the ratio between the number of land registry parcels with ghost buildings and the total number of land registry parcels. **Extra Controls*Post** include *Geographic*, *Socio-Economic*, and *Mayor Controls* interacted with the *Post* dummy. Refer to Table 2.1 for a list of these variables. All the regressions include town fixed effects, election-year fixed effects and an interaction between macro-areas fixed effects and *Post*. Standard errors are clustered at provincial level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

potential explanations about the impact of the Ghost Buildings program on voter support for the local incumbent.

According to the first of these alternative explanations, the publication of the number of ghost buildings generates information about the incumbent. We believe this to be both unlikely and inconsistent with our findings. First, the set of ghost buildings is a slow moving stock variable that is likely to have accumulated over decades, rather than a reflection of just the most recent years. Most of the buildings found by the *Agenzia del Territorio* were not newly constructed. The existence of a term limit, paired with the fact that the time to complete a building in Italy is generally longer than most of the other OECD countries, suggests that most of these buildings could not have been built while the incumbent was in office. Second, we notice that voters who could potentially receive information from the publication are most likely the ones who were not evading before the program, as evaders already knew about their own evasion.

Keeping this premise in mind, our results rule out this alternative explanation. In one version of this alternative story, voters, after learning about *low* levels of evasion detected by

the program, reward the current mayor for having properly enforced tax payment in the past. This hypothesis predicts a *negative* impact of the detected ghost building intensity on incumbent reelection in post-program elections, and as such it is obviously inconsistent with our baseline results. In another version of this alternative explanation, voters reward an incumbent mayor for having allowed *high* levels of evasion in the past. First, this contradicts the intuition that non evaders, rather than those previously evading, are the ones who are potentially acquiring new information. Second, it is unlikely since the purpose of the program, and therefore the publication, was to shut down the evasion opportunity. Third, it is at odds with the fact that the positive impact of program intensity on incumbent reelection is lower in regions with higher tolerance for tax evasion. Fourth, it is inconsistent with our results showing that towns with higher registration levels are more likely, rather than less likely, to reelect an incumbent mayor.

In a second potential alternative explanation, the program gives an incumbent an electoral rent by allowing her *to not register* the targeted ghost buildings, for instance by reporting errors in the results generated by the mapping process.⁴⁷ If this were the case, we would expect the positive impact of the program to be stronger in regions with higher tolerance for tax evasion. We find the opposite to be the case. In addition, such an explanation is inconsistent with the result that a higher share of registered ghost buildings at the time of a local election increases reelection likelihood. The results of this section provide strong evidence that it is the additional tax enforcement induced by the program that drives the increase in the reelection prospects of the incumbent.

2.7 Conclusion

A rapidly growing literature shows that interventions that improve the “technology” of tax enforcement — third party reporting, cross-checking, or better auditing algorithms — can

⁴⁷For example, the press agency of one of the mayor of a city in our sample, Capaccio Paestum, explicitly criticized the “excessive media attention” to the program, indicating how the unregistered buildings in that city were due to “citizens’ needs” (Comune di Capaccio Paestum, 2010).

substantially reduce tax evasion. Yet, political incentives to adopt these technologies are also of crucial importance. Policymakers will delay or prevent enforcement policies if they are bound to lose support from them. In spite of this, little is known about the electoral impact of fighting tax evasion. This paper provides evidence of a positive interaction between technological improvements in tax payer monitoring and political incentives. Specifically, local incumbents are shown to obtain positive political returns — an increase in their reelection likelihood — from the Ghost Building program, a nationwide anti tax evasion policy in Italy which was based on a new enforcement technology.

Underlying tax culture — broadly defined as the individual propensity and social norms determining evasion for a given level of technology — is another important determinant of tax compliance. It shapes the enforcement level a government can achieve for a given enforcement technology. We show that tax culture affects the political returns to undertaking anti tax evasion policies. The increase in incumbent reelection probabilities in response to the Ghost Buildings program is larger in areas with lower self-reported tolerance for tax evasion. Finally, we document that the political returns to enforcement policies are higher when the government is more efficient in public goods provision.

The findings of the paper have two important policy implications. First, they provide a framework for thinking about the political feasibility of policies that increase visibility of tax evasion, thus lowering the monitoring costs and increasing policymakers' incentives for raising enforcement. This has immediate relevance for special interest politics. Concentrated evader groups might effectively lobby to keep evasion hidden from the public. Yet, they are unlikely to be able to punish an incumbent who enforces tax compliance after the evasion becomes broadly visible.

Second, there is potential complementarity among anti tax evasion policies, government responsiveness, and social preferences for tax compliance. Governments that plan to implement novel enforcement policies should concurrently attempt to strengthen their capabilities, for instance by improving the speed at which they respond to citizen's needs, or by increasing social stigma associated with tax evasion. This complementarity will likely

increase the returns politicians obtain from anti tax evasion policies and will thus make such policies more aligned with political agent incentives.

We are aware that using an identification strategy based on a specific natural experiment enhances internal validity of our study but might come at the price of lower external validity, for instance for extrapolating about similar programs in other countries or programs targeting other taxes. Yet, we speculate that evidence of positive political returns to anti tax evasion policies in Italy, a country often cited as an example of poor tax culture, will be a lower bound for other OECD countries. We believe an interesting goal for future work is to elucidate the potential non-linearity in the relation between tax evasion prevalence and political returns to enforcement policies. In addition, we believe that complementarity between enforcement policies and social norms on evasion could potentially be relevant for policy design in other regions of the world.

Another important dimension of external validity concerns enforcement policies targeting other types of evasion. One of the merits of the Ghost Buildings program is that it detected the entire stock of evasion. On the other hand, the effectiveness of policies targeting other tax-concealing activities might vary by the ability of the specific evader to hide. This might affect how the public would respond. We hope future work will shed light on the political returns to other enforcement policies around the world.

Chapter 3

Law, Economics and Culture: Theory of Mandated Benefits and Evidence from Maternal Leave Policies¹

3.1 Introduction

Mandated maternity leave is one of the key policies that supports continued labor force participation of women. However, there is a large variation in the length of maternity leave that countries mandate. Why do some countries mandate that employers provide a long maternity leave, while others mandate only a short one?

To address this question we present a mandated benefit model, with an employer, and two types of employees, men and women. The employer is required to provide maternity leave to women and not to men. The employer's cost of providing maternity leave is greater than its value to women, since otherwise there would be no need for the mandate as the employer would freely choose to provide the leave (Summers 1989). Therefore, when the employer is free to discriminate between men and women in their wages, women's wage and employment goes down due to the mandate.

¹Co-authored with Yehonatan Givati.

We then incorporate into the model society's attitudes towards gender based discrimination. In a society where there is some negative view of gender based discrimination the employer is not free to treat men and women differently. This could be the result of a social sanction imposed on employers that treat men and women too differently, or because of a legal provision that restricts the extent of permissible discrimination. Formally, the less willing society is to tolerate gender based discrimination the less able the employer is to discriminate between men and women in their wages.

When gender based wage discrimination is restricted, mandated maternity leave can benefit women, as shown by Jolls (2000, 2006), since men bear some of the cost of the mandated leave. We show that the less society is willing to tolerate gender based discrimination the longer the maternity leave it will provide. Intuitively, when society is relatively intolerant of gender based discrimination employers are less able to pass on to women the cost of maternity leave. This means that an increase in the length of maternity leave has a smaller effect on women's wages, and therefore a longer leave will benefit them.

To see whether the model's predictions are supported by the data, we use data on length of maternity leave in different countries, as well as other economic, political and demographic controls. Looking for a measure of society's tolerance of gender based discrimination, we begin by using answers to the World Value Survey. In a cross sectional analysis we show that countries where gender based discriminatory views are more prevalent have shorter maternity leave, which is consistent with the prediction of our model.

However, because answers to surveys may be endogenous to current policies, we look for a measure of society's attitudes towards gender based discrimination for which concerns of reverse causation are mitigated. To do so we rely on recent research in psychology and linguistics according to which language shapes a person's view of the world. This view, known as the Sapir-Whorf hypothesis of linguistic relativity, originally advanced by Edward Sapir (1929) and Benjamin Lee Whorf (1956), is well summarized by the psychologist Lera Boroditsky (2010):

Patterns in language offer a window on a culture's dispositions and priorities. . .

new research shows us that the languages we speak not only reflect or express our thoughts, but also shape the very thoughts we wish to express. . . As we uncover how languages and their speakers differ from one another, we discover that human nature too can differ dramatically, depending on the languages we speak.

The stable grammatical feature we focus on is personal pronouns. Languages differ in how many of their personal pronouns are gender differentiated. We collected data on the number cases of gender differentiated pronouns in 33 languages. To the best of our knowledge, this is the first time that this variable has been systematically coded and used in an empirical quantitative work. Using individual answers to the World Values Survey, while exploiting only within country variation, we shows that speakers of languages with more cases of gender differentiated personal pronouns are more likely to have gender based discriminatory views. Intuitively, a language that routinely compels you to specify gender when using personal pronouns increases your awareness and acceptance of gender differences. This is consistent with the linguistic relativity principle, and supports our use of this feature of languages as a proxy for attitudes towards gender based discrimination.

Using our new linguistic measure in a cross sectional analysis, we find that the more cases of gender differentiated personal pronouns a language that is spoken in a country has, the shorter the maternity leave that a country provides, which is consistent with the prediction of our model.

The question why countries vary in their mandated maternity leave policies has been of central concern to legal scholars. In particular, the focus has been on explaining why the U.S. mandates only limited maternity benefits, as opposed to the more expansive benefits mandated in certain European countries (Dowd 1989, Issacharoff and Rosenblum 1994, Schuchmann 1995, Grill 1996, Pelletier 2006, Levmore 2007, Suk 2010). The reasons for the different policies mentioned in these papers include: different approaches to the role of the state (European countries being social welfare states, and the U.S. being individualistic and market oriented), different demographic needs (low fertility in Europe relative to the U.S.), differing goals of the feminist movement (striving for equal treatment in the U.S., and for special treatment in Europe), and different legal structures for providing benefits (in

the U.S. antidiscrimination law is relied upon, which places maternity and medical leave under the same legal regime, while in Europe maternity leave is covered by special laws). In our empirical analysis we control, where possible, for these alternative explanations. More importantly, we expand the analysis significantly, and instead of comparing the U.S. with a couple of European countries we utilize data on maternity leave policies and their possible determinants in around 80 countries. The advantages of large sample analysis when addressing comparative law questions are well noted in Spamann (2009).

That maternity leave laws are determined endogenously is in the spirit of Aghion, Alesina and Trebbi (2004), who analyze the endogenous choice of political institutions. Along the same lines Aghion, Algan, Cahuc and Shleifer (2010) analyze the feedback between regulation and distrust, Nannicini, Stella, Tabellini and Troiano (2010) analyze the relation between social capital and political accountability, and Givati (2011) analyzes the endogenous determination of plea bargaining regimes.² The paper is also related to the literature on the effects of culture on economic policies and outcomes (Putnam 1993, 2000; Fukuyama 1995; Bisin and Verdier 2000, 2001; Guiso, Sapienza, and Zingales 2006, 2008, 2009; Tabellini 2008; Alesina and Giuliano 2010).

(Putnam 1993, 2000; Fukuyama 1995; Guiso, Sapienza, and Zingales 2006, 2008, 2009; Tabellini 2008).

This part proceeds as follows. Section 3.2 describes the model. Section 3.3 lays out our empirical strategy. Section 3.4 shows that accounting for paternity leave and multilingual countries does not affect our results, and that they are not driven by countries where Arabic is spoken. It also discusses the possibility of an alternative channel that is consistent with our results. Section 3.5 concludes.

²For papers that look at the effect of maternity leave laws, see Ruhm (1998), Ruhm (2000) and Lalive and Zweimüller (2009).

3.2 The Model

3.2.1 Setup

An employer uses labor (L) for production, with the following production function:

$$F(L) = ZL - \frac{1}{2}L^2 \quad (3.1)$$

where Z is some large parameter. The price of the good produced is normalized to equal one.

The population in the economy consists of a unit mass of men and a unit mass of women. Because of government regulation the employer is legally required to provide maternity leave of length μ to women. This imposes a cost $c(\mu)$ on the employer, when employing women, and its value to women is $v(\mu)$. The utility functions of men and women, U_m and U_w respectively, are:

$$\begin{aligned} U_m &= W_m L_m - \frac{1}{2} L_m^2 \\ U_w &= (W_w + v(\mu)) L_w - \frac{1}{2} L_w^2 \end{aligned} \quad (3.2)$$

where W_m and L_m are the wage and labor of men, and W_w and L_w are the wage and labor of women.

We assume that $c(\mu) > v(\mu)$, that is the employer's cost of providing maternity leave is greater than its value to women. If this was not the case there would be no need for the mandate, as the employer would freely choose to provide the leave to women absent government regulation (Summers 1989). Assume also that $c'(\mu), v'(\mu) > 0$, and that $c''(\mu) \geq 0 > v''(\mu)$, that is women's marginal benefit of maternity leave is positive and decreasing, while the employer's marginal cost of maternity is positive and non-decreasing.

3.2.2 Gender Based Discrimination

We first analyze the case where the employer is free to offer different wages to men and women. Using expression 3.2 we can derive men's and women's labor supply:

$$L_m = W_m \quad (3.3)$$

$$L_w = W_w + v(\mu)$$

Intuitively, men's labor supply is increasing with their wages, and women's labor supply is increasing with their wages and with their valuation of maternity leave.

The employer solves a separate maximization problem for men and women, taking in each case the number of workers from the opposite gender as given:

$$\max_{L_m} F(L_m + \bar{L}_w) - W_m L_m$$

$$\max_{L_w} F(\bar{L}_m + L_w) - (W_w + c(\mu)) L_w$$

These yield the employer's demand function for men and for women:

$$W_m = Z - L_m - \bar{L}_w \quad (3.4)$$

$$W_w = Z - \bar{L}_m - L_w - c(\mu)$$

Naturally, the demand curve for men is decreasing with the number of men employed, and the demand curve for women is decreasing with the number of women employed. Note that an increase in the number of women employed shifts the demand curve for men downward, and an increase in the number of men employed or in the cost of providing maternity leave shifts the demand curve for women downward.

We can now plug in the labor supply functions into the demand functions, and solve for the wage of men and women:

$$W_m^* = \frac{1}{3}[Z - v(\mu) + c(\mu)] \quad (3.5)$$

$$W_w^* = \frac{1}{3}[Z - v(\mu) - 2c(\mu)]$$

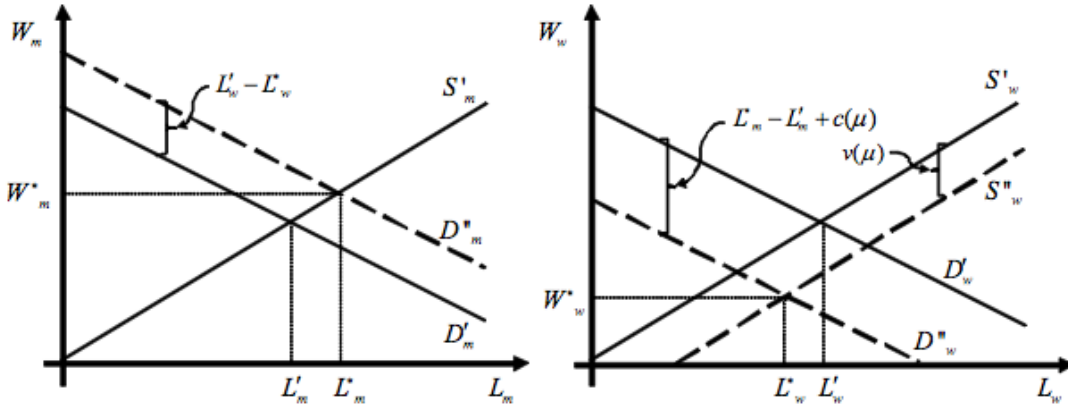
The number of men and women employed, L_m^* and L_w^* respectively, can be immediately derived from the labor supply functions in expression 3.3.

Note from expression 3.5 that $W_m^* - W_w^* = c(\mu)$, that is the difference in wages between men and women is equal to the cost of providing maternity leave. Intuitively, since relative to men employing women imposes a cost $c(\mu)$ on the employer, women's wage is lower than men's wage by this cost.

Let us denote the equilibrium wages and employment before maternity leave is mandated as W'_m, W'_w, L'_m, L'_w . It is easy to see that when no maternity leave is mandated, that is when $c(\mu) = v(\mu) = 0$, we get $W'_m = W'_w = L'_m = L'_w = \frac{1}{3}Z$. Note that for men $W_m^* > W'_m$ and $L_m^* > L'_m$, and for women $W_w^* < W'_w$ and $L_w^* < L'_w$. That is, as a result of the mandated leave women's wage and employment goes down, whereas men's wage and employment goes up.

Figure 3.1 depicts the labor market for men and women.

Figure 3.1: *The Labor Market for Men (left) and Women (right)*



The introduction of mandated maternity leave shifts women's supply curve outward by $v(\mu)$, and the demand curve for women inward by $c(\mu)$. Since $c(\mu) > v(\mu)$ this results in a decrease in employment of women, which in turn shifts the demand curve for men outward, as expression 3.4 shows. The resulting increase in employment of men causes a further shift inward in the demand curve for women.

3.2.3 Intolerance of Gender Based Discrimination

We now analyze how society's tolerance of gender based discrimination affects the analysis presented above. In a society where there is some negative view of gender based discrimination the employer is not free to treat men and women differently. This could be the result of a social sanction imposed on employers that treat men and women too differently, or because of a legal provision that restricts the extent of permissible discrimination.

To capture society's attitudes towards gender based discrimination formally, we impose the following condition:

$$W_w \geq W_m - \frac{1}{\lambda}c(\mu) \quad (3.6)$$

where $\lambda \geq 1$. Recall that $W_m^* - W_w^* = c(\mu)$. λ reflects an external constraint on gender based wage discrimination, and therefore captures how much society is willing to tolerate gender based discrimination. The less willing society is to tolerate gender based discrimination the higher λ is, and therefore the smaller the difference between the wage of men and women. To make the constraint in expression 3.6 meaningful, we assume that the employer must hire women and men willing to work for the wages offered.

Jolls (2000, 2006) also analyzes the effect of mandated benefit that is targeted at a specific population, when employers cannot discriminate in wages between the different populations. Here, we employ a similar framework, but instead of imposing that the wage of men and women must be equal, we consider the extent to which the wage of men and women can be different, and analyze the consequences of a change in the constraint on wage differentiation on the level of mandated benefits.

Plugging in the labor supply functions from expression 3.3 for L_m and L_w in the demand functions in expression 3.4 yields the following two conditions:

$$\begin{aligned} W_m &= \frac{1}{2}[Z - \bar{L}_w] \\ W_w &= \frac{1}{2}[Z - \bar{L}_m - v(\mu) - c(\mu)] \end{aligned} \quad (3.7)$$

Imposing the constraint from expression 3.6 on the conditions in expression 3.7, and solving

for the wage of men and women, we get:

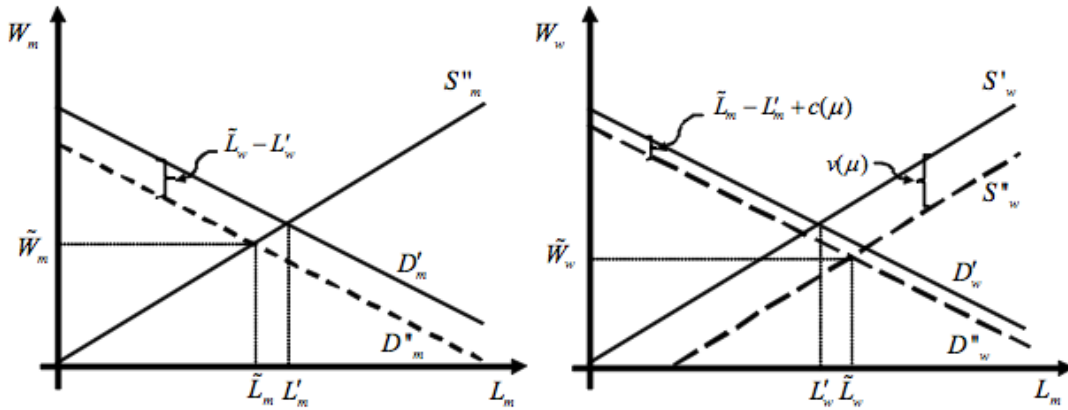
$$\begin{aligned}\tilde{W}_m &= \frac{1}{3}[Z - v(\mu) + (\frac{2-\lambda}{\lambda})c(\mu)] \\ \tilde{W}_w &= \frac{1}{3}[Z - v(\mu) - (\frac{\lambda+1}{\lambda})c(\mu)]\end{aligned}\tag{3.8}$$

The number of men and women employed, \tilde{L}_m and \tilde{L}_w respectively, can again be immediately derived from the labor supply functions in expression 3.3.

Note that when $\lambda = 1$, that is when society is indifferent to gender based discrimination, we get $\tilde{W}_m = W_m^*$ and $\tilde{W}_w = W_w^*$. Also, note that $\frac{\partial \tilde{W}_m}{\partial \lambda} < 0$, while $\frac{\partial \tilde{W}_w}{\partial \lambda} > 0$, that is as society becomes less tolerant of gender based discrimination men's wage goes down, while women's wage goes up.

Figure 3.2 depicts the labor market for men and women in a society that is not indifferent to gender based discrimination, that is when $\lambda > 1$.

Figure 3.2: *The Labor Market for Men (left) and Women (right) when Society is Intolerant of Discrimination*



In equilibrium, the shift inward of the demand for women by $c(\mu)$ as a result of the introduction of mandated maternity leave is offset by a shift outward of the same curve as a result of a decrease in the employment of men. The increase in employment of women shifts the demand curve for men inward. One can see that, relative to the case where there is no mandate, men's wage and employment go down, women's employment goes up while

their wage goes down.

3.2.4 Optimal Length of Leave

A social planner maximizes social welfare, which depends on the utility of men and women, $SW = SW(U_m, U_w)$. Following Diamond (1975), let us define the social marginal utility of men's and women's income, $\frac{\partial SW}{\partial U_m} = \beta_m$, and $\frac{\partial SW}{\partial U_w} = 1$. Assume that $\beta_m < 1$, which means that the social marginal utility of women's income is greater than that of men's. This reflects a case where men's utility is higher than women's, and the social welfare function is concave, expressing an aversion to inequality in the distribution of utility levels between men and women. The social planner chooses the length of maternity leave to maximize social welfare.

To simplify the analysis, and develop the basic intuition, let us assume that $\beta_m = 0$ (Appendix C.1 presents the full analysis where $\beta_m > 0$).³ Differentiating the social welfare function with respect to the length of leave, μ , we employ the envelope theorem and ignore $\frac{\partial L_w}{\partial \mu}$, as women choose their labor supply optimally. We thus get:

$$\frac{\partial SW}{\partial \mu} = \left(\frac{\partial W_w}{\partial \mu} + v'(\mu) \right) L_w$$

Using women's wage from expression 3.8, we obtain the following first order condition:

$$v'(\mu) = \left(\frac{1 + \lambda}{2\lambda} \right) c'(\mu) \quad (3.9)$$

The length of maternity leave for which expression 3.9 hold is denoted by $\tilde{\mu}$.⁴ Intuitively, increasing the length of maternity leave has a direct beneficial effect on women, but it also results in reduction in women's wage, which harms women. The left-hand side of expression 3.9 is women's direct gain from increasing the length of maternity leave. The right-hand side of expression 3.9 is the women's loss from increasing the length of maternity leave, since it is the increase in the cost imposed on the employer when employing women which

³This assumption could also reflect a situation where maternity leave is introduced only to benefit women, perhaps due to political considerations.

⁴This is a maximum, since the second order condition holds: $v''(\mu) - \left(\frac{1 + \lambda}{2\lambda} \right) c''(\mu) < 0$.

reduces women's wages, attenuated by society's intolerance of gender based discrimination. At the optimum the marginal gain and marginal loss from a longer maternity leave are equal.

3.2.5 The Effect of Attitudes on Leave

How do society's attitudes towards gender based discrimination affect the length of the maternity leave mandated? To address this question we look at the effect of λ , which captures society's tolerance of gender based discrimination, on the chosen length of maternity leave, $\tilde{\mu}$.

Employing the implicit function theorem on the first order condition in expression 3.9, we get:

$$\frac{\partial \tilde{\mu}}{\partial \lambda} = \frac{c'(\mu)}{2\lambda^2[(\frac{1+\lambda}{2\lambda})c''(\mu) - v''(\mu)]} > 0 \quad (3.10)$$

That is, the less society is willing to tolerate gender based discrimination the longer the maternity leave it will provide. Intuitively, when society is relatively intolerant of gender based discrimination employers are less able to pass on to women the cost of maternity leave. This means that an increase in the length of maternity leave has a smaller effect on women's wages, and therefore a longer leave can be chosen.

3.3 Empirical Analysis

3.3.1 Cross Country Correlation

As section 3.2 explains, we are interested in the relationship between society's tolerance of gender based discrimination and the length of maternity leave it provides. To look at attitudes towards gender based discrimination, we use data collected in the World Value Survey (see appendix C.2 for source information). In particular, we look at respondents' answer to the following question: "When jobs are scarce, men should have more right to a job than women? Agree/ Disagree". For each country we use the percentage of respondents who thought that when jobs are scarce men should have more right to a job than women,

Table 3.1: *Summary Statistics for Maternity Leave and Attitudes Regression*

	Mean	S.D.	Min	Max
Days of Maternity Leave	137.1	87.09	60	480
Men Have More Right to a Job	46.62	23.27	5	100
GDP per capita	18,582	14,752	886	79,813
Govt. Share of GDP	16.25	5.9	6.32	36.54
Population (in thousands)	77,254	212,942	329	1,354,146
Fertility	2.03	0.86	1.2	5.9
Women Share of Parliament	18.26	10.54	0	45

averaged over the waves of the World Value Survey.⁵

In a cross sectional analysis, we estimate the correlation between the length of maternity leave and the percentage of respondents in a country who think men should have more right to a job than women, which we interpret as capturing attitudes towards gender based discrimination. The data on the length of maternity leave in different countries, as well as other demographic and political variable we use, are from the United Nations Statistical Division, Statistics and Indicators on Women and Men. GDP data is taken from the Penn World Table (see appendix C.2 for source information). Summary statistics for the variables we include in our regression are presented in table 3.1.

From table 3.1 it is clear that there is a large variation in the length of maternity leave that countries mandate, from 60 days (Uganda and the Philippines) to 480 days (Sweden). Similarly, there is a large variation in the percentage of respondents in each country who think that men should have more right to a job than women.

Table 3.2 presents the results of the OLS regression. The dependent variable in Table 3.2 is the number of days of maternity leave a country mandates. Standard errors are robust.

In specification (1) we can see that there is a strong negative correlation between the length of maternity leave and the percentage of respondents who think men should have

⁵We calculated the data for every country and wave of survey using the World Value Survey online analysis, to avoid any risk of manipulation of the raw data. Then an average was taken for every country over the four first waves of the World Value Survey. The fifth wave was not included here since it was not available for online analysis.

Table 3.2: Maternity Leave and Attitudes

Dependent Variable:	Days of Maternity Leave			
	(1)	(2)	(3)	(4)
Men Have More Right to a Job	-1.39 (0.46)***	-1.21 (0.50)**	-0.922 (0.428)**	-0.80 (0.37)**
GDP per capita		0.0006 (0.0009)	0.0005 (0.0009)	0.0001 (0.0008)
Government Share of GDP		2.36 (1.55)	2.27 (1.49)	3.69 (1.51)**
Population		-0.00005 (0.00002)***	-0.00005 (0.00002)***	-0.00005 (0.00001)***
Women Share of Parliament			1.22 (1.17)	1.59 (1.17)
Fertility				-30.51 (9.48)***
R^2	0.14	0.17	0.18	0.25
Number of obs.	73	73	73	73

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Standard errors are robust

more right to a job than women. Specifically, for every additional percent of respondents who think men should have more right to a job than women the length of maternity leave decreases by 1.4 days.

In specification (2) we control for GDP per capita, government share of GDP and population. GDP per capita is not correlated with the length of maternity leave. Government share of GDP, which can be thought of as a proxy for the size of the welfare state, is positively correlated with the length of maternity leave, but this correlation is not statistically significant. The population of country is negatively correlated with the length of maternity leave. The effect of attitudes on leave remains statistically and economically significant after adding these controls.

In Specification (3) we control for the share of women in parliament. The coefficient is positive, which may reflect a political influence on length of leave, but the coefficient is not statistically significant. The effect of attitudes on leave remains statistically and economically significant.

In specification (4) we control for fertility rate. The coefficient is negative and statistically

significant, that is a lower fertility rate is associated with a longer leave, which could reflect a policy intended to increase fertility by providing maternity leave. The effect of attitudes on leave remains statistically and economically significant even after controlling for fertility.

In all specifications in table 3.2 we see that there is a strong negative correlation between the length of maternity leave and gender based discriminatory views, and this effect is statistically and economically significant. That is, the more people in a country think that when jobs are scarce men should have more right to a job than women, the shorter the maternity leave that a country mandates. This is consistent with the prediction of our model, that the more tolerant a society is towards gender based discrimination the shorter the maternity leave it mandates.

3.3.2 Language and Attitudes

Linguistic Relativity

As subsection 3.3.1 shows, there is a strong correlation between the length of maternity leave and gender based discriminatory views. However, the main challenge in looking at the effect of attitudes on policies is that current measure of attitudes through surveys may be influenced by current policies. In our case, a maternity leave policy that is favorable to women might produce survey answers that are favorable to women. Ideally, to provide a correlation between long run attitudes and current policies we would like to have a survey done centuries ago and estimate the correlation. We try to deal with this issue using linguistic relativity.

Linguistic relativity is the idea that thought is shaped by language, or more precisely that the particular language we speak influences the way we think about reality. This notion was first advanced by Edward Sapir (1929), but was mostly developed in the writing of his student, Benjamin Lee Whorf (1956). The main thrust of what is known as the Sapir-Whorf hypothesis is well summarized by the linguist Guy Deutscher (2010a):

When your language routinely obliges you to specify certain types of information, it forces you to be attentive to certain details in the world and to certain aspects

of experience that speakers of other languages may not be required to think about all the time. And since such habits of speech are cultivated from the earliest age, it is only natural that they can settle into habits of mind that go beyond language itself, affecting your experiences, perceptions, associations, feelings, memories and orientation in the world.

Although the Sapir-Whorf hypothesis was rejected by scientists for many years due to lack of supporting evidence, in the past fifteen years a solid body of empirical work has emerged to support it (see surveys of the literature in Gumperz and Levinson 1996, Lucy 1997, Boroditsky 2003, Deutscher 2010b, Boroditsky 2011). Studies in psychology and linguistics, each focusing on a specific cross-linguistic difference, have confirmed the effect of language on thought. For example, speakers of languages that use cardinal-direction terms – north, south, east, and west – to define space, instead of defining space relative to an observer, have superior navigational ability and spatial knowledge, and apply the same frame of reference in recall and recognition (Levinson 2003). The way languages divide the color spectrum into colors leads to differences in color discrimination (Winawer et al. 2007). Speakers of languages that do not make significant grammatical distinction between objects and substances (for example, when counting them), pay more attention to the material of objects rather than to their shape (Lucy and Gaskins 2001).

Applying the linguistic relativity principle to our research, we look for a grammatical feature of languages that is correlated with attitudes towards gender based discrimination.

Gender Differentiated Personal Pronouns

One of the stable grammatical features of a language is its use of personal pronouns. Personal pronouns are gender differentiated in some cases, but in other cases they are not. Languages differ in how many of their personal pronouns are gender differentiated. For example, as shown in table 3.3, Spanish has four cases of gender differentiated pronouns (third person singular, first person plural, second person plural, and third person plural), whereas English has only one case of gender differentiated pronoun (third person singular).

We code, using grammar books, the number of cases of gender differentiated personal

Table 3.3: *Personal Pronouns in English and Spanish*

Person	English				Spanish			
	Singular		Plural		Singular		Plural	
	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
1st	I		We		yo		nosotros	nosotras
2nd	You		You		tú		vosotros	vosotras
3rd	He	She	They		él	ella	ellos	ellas

pronouns in 33 languages (see data in appendix C.3). To the best of our knowledge, this is the first time that this variable has been systematically coded and used in an empirical quantitative work.⁶

Applying the linguistic relativity principle to our research, a language that routinely compels you to specify gender when using personal pronouns increases your awareness and acceptance of gender differences. Thus, the number of gender differentiated personal pronouns can be used as a proxy for attitudes towards gender based discrimination, where the fewer gender differentiated personal pronouns a language has the less tolerant its speakers should be towards gender based discrimination.

We are not the first to look at the effects of variation in grammatical gender across languages. Guiora et al. (1982) study the development of gender identity in children who speak Hebrew, English and Finnish, as the sex-determined grammatical features of these languages, and in particular their use of gender differentiated pronouns, vary from almost zero in Finnish, through very low in English, to very high in Hebrew. They find a direct relationship between gender loading in the native language and gender identity attainment. Hill and Mannheim (1992) note the "Whorfian effect" of gender differentiated personal pronouns, and more recently, Boroditsky et al. (2003) find that the way languages assign grammatical gender to inanimate objects affects how their speakers view these objects.

In the economics literature, other features of language are used as a proxy for culture by Licht, Goldschmidt and Schwarz (2006) and Tabellini (2008).

⁶We selected the languages we used based on how widely they were spoken, and the availability of grammar books for us to rely on. We estimate that the languages we currently have cover approximately 53% of world population.

Is Language Correlated with Attitudes?

Before turning to aggregate cross-country data, it is natural to ask whether our linguistic variable is correlated with attitudes toward gender based discrimination. To do so we use again data collected in the World Value Survey. In particular, we look at individual answers to the question: "When jobs are scarce, men should have more right to a job than women? Agree/ Disagree".⁷ One of the advantages of the World Value Survey data is that it allows us distinguish between the country where the respondent lives (from the question "In which country do you live?") and the language the respondent speaks (from the question: "What language do you normally speak at home?"). Every respondent was assigned the number of cases of gender differentiated pronouns the language that he reported to be speaking at home has. If indeed language is correlated with attitudes, as we argue, then we should find that the number of cases of gender differentiated pronouns is associated with a higher likelihood of agreeing with the above statement on men's right to a job.

Summary statistics for the variables we include in our regression, all taken from the World Value Survey, are presented in table 3.4

Table 3.5 presents the results. The dependent variable in Table 3.5 is a binary variable representing the person's answer to the question whether men have more right to job than women when jobs are scarce. The variable takes the value 1 if the respondent agreed with the statement, and zero if the respondent disagreed with it. The table presents marginal effects of a probit regression. Standard errors are robust and clustered at the language level, to allow for arbitrary patterns of correlation by language. We include controls for country and year of survey fixed effects. Thus, we only exploit the within country and within year variation in responses, which means we hold constant policies, institutions and laws that might also have an impact on individual attitudes.

⁷We use this question because it is directly related to gender based discrimination in the workforce, which is what our model is about, and because it seems to closely reflect attitudes towards gender based discrimination more than any other question on the World Value Survey. As a robustness check, we looked at other gender related questions in the World Value Survey, and find similar results for questions on whether men make better political leaders than women, whether a woman has to have children to be fulfilled, and whether a wife must always obey her husband.

Table 3.4: *Summary Statistics for World Value Survey Regression*

	Mean	S.D.	Min	Max
When jobs are scarce men should have more right to a job than women	0.39	0.49	0	1
Gender Differentiated Personal Pronouns	2.14	1.51	0	4
Male	0.48	0.50	0	1
Married	0.60	0.49	0	1
High School	0.34	0.47	0	1
College or more	0.16	0.36	0	1
Income	4.60	2.47	1	10
Age	41.4	16.2	15	98

In specification (1) we see that the more gender differentiated pronouns a language has the more likely the respondent is to think that when jobs are scarce men have more right to a job than women, and this effect is statistically significant.

In specification (2) we control for the gender of the respondent. Unsurprisingly, men are more likely to think that when jobs are scarce men have more right to a job than women. However, even after controlling for the gender of the respondent, the effect of language on the respondents answer remains significant.

In specification (3) we see that older people and people who are married are more likely to agree with the statement, while people with higher income are less likely to do so. Still, the effect of language on the respondents answer remains significant.

In specification (4) we see that the more educated a person is the less likely the person is to agree with the statement. Still, even after controlling for education, the more gender differentiated pronouns a language has the more likely the respondent is to think that when jobs are scarce men have more right to a job than women.

That speakers of languages with more cases of gender differentiated personal pronouns are more likely to have gender based discriminatory views is consistent with the linguistic relativity principle, and supports our use of this feature of languages as a proxy for attitudes towards gender based discrimination.

Table 3.5: Language and Attitudes

Dependent Variable:	When jobs are scarce men should have more right to a job than women			
	(1)	(2)	(3)	(4)
Gender Differentiated	0.018	0.019	0.023	0.022
Personal Pronouns	(0.008)**	(0.008)**	(0.008)***	(0.008)***
Male		0.103	0.108	0.113
		(0.013)***	(0.011)***	(0.012)***
Age			0.0028	0.0024
			(0.0006)***	(0.0006)***
Married			0.045	0.046
			(0.013)***	(0.012)***
Income			-0.021	-0.015
			(0.003)***	(0.003)***
High School				-0.067
				(0.007)***
College or more				-0.145
				(0.011)***
Country FE	✓	✓	✓	✓
Survey Year FE	✓	✓	✓	✓
Number of Obs.	96,233	96,233	96,233	96,233

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Standard errors are robust and clustered at the language level

3.3.3 Maternity Leave and Language

In a cross sectional analysis, we estimate the correlation between the length of maternity leave and gender differentiated personal pronouns, interpreted as a proxy for attitudes towards gender based discrimination. In addition to the data that was used before, in subsection 3.3.1, here we associate each country in the sample with the language most commonly spoken in it, based on the CIA World Factbook and Ethnologue (see appendix C.2 for source information). Summary statistics for the variables we include in our regression are presented in Table 3.6.

From Table 3.6 it is clear that there is a large variation in the length of maternity leave that countries mandate. In our sample, the length of maternity leave goes from 45 days (Bahrain and UAE) to 480 days (Sweden). Similarly, there is a large variation in the number of gender differentiated personal pronouns of languages spoken in the countries in our

Table 3.6: *Summary Statistics for Maternity Leave and Language Regression*

	Mean	S.D.	Min	Max
Days of Maternity Leave	118.0	78.12	45	480
Gender Differentiated Personal Pronouns	2.49	1.51	0	4
GDP per capita	18,837	14,828	1,309	79,813
Govt. Share of GDP	15.98	6.75	5.8	37.97
Population (in thousands)	60,616	202,116	257	1,354,146
Fertility	2.21	0.94	1.3	5.8
Women Share of Parliament	16.84	10.7	0	45

sample.

Table 3.7 presents the results of the OLS regression, which we consider our main results. The dependent variable in Table 3.7 is the number of days of maternity leave a country mandates. Standard errors are robust and clustered at the language level.

In specification (1) we can see that there is a strong negative correlation between the length of maternity leave and our linguistic measure. The coefficient is negative, statistically and economically significant. Specifically, for every additional case of gender differentiated pronouns the length of maternity leave decreases by approximately 20 days.

In specification (2) we control for GDP per capita, government share of GDP and population. GDP per capita is not correlated with the length of maternity leave. Government share of GDP, which can be thought of as a proxy for the size of the welfare state, is positively correlated with the length of maternity leave, but this correlation is not statistically significant. The population of country is negatively correlated with the length of maternity leave. The effect of language on leave remains statistically and economically significant after adding these controls.

In Specification (3) we control for the share of women in parliament. The coefficient is positive and statistically significant. Specifically, an increase of one percent in women's representation in parliament is associated with an increase in the length of mandated maternity leave of 2 days. This results seems intuitive, and captures the political channel

Table 3.7: Maternity Leave and Language

Dependent Variable:	Days of Maternity Leave			
	(1)	(2)	(3)	(4)
Gender Differentiated	−19.95	−21.88	−18.61	−16.78
Personal Pronouns	(5.72) ^{***}	(7.00) ^{***}	(4.00) ^{***}	(4.22) ^{***}
GDP per capita		0.0005	0.0001	0.0001
		(0.0008)	(0.0005)	(0.0005)
Government		1.80	1.53	1.92
Share of GDP		(1.40)	(0.99)	(1.09) [*]
Population		−0.00008	−0.00007	−0.00007
		(0.00002) ^{***}	(0.00002) ^{***}	(0.00001) ^{***}
Women Share			2.06	1.94
of Parliament			(1.08) [*]	(1.12) [*]
Fertility				−8.85
				(6.62)
R^2	0.15	0.21	0.27	0.28
Number of obs.	81	81	81	81

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Standard errors are robust and clustered at the language level

that affects the length of maternity leave. However, even after controlling for this political channel the effect of language on leave remains statistically and economically significant.

In specification (4) we control for fertility rate. A lower fertility rate is associated with a longer leave, which could reflect a policy intended to increase fertility by providing maternity leave, but this effect is not statistically significant. The effect of language on leave remains statistically and economically significant even after controlling for fertility.

The results shown in Table 3.7 are consistent with our model, that is with the prediction that lower tolerance of gender based discrimination, which we capture here through our language proxy, is associated with a longer mandated maternity leave.

3.3.4 Instrumental Variable Approach

If language structure, that is the number of gender differentiated personal pronouns, affects the length of maternity leave only through its effect on attitudes towards gender based discrimination, one can employ an instrumental variable approach to look at the effect of

Table 3.8: Instrumental Variable Approach

Dependent Variable:	(1) & (3):	When jobs are scarce men should have more right to a job than women		
	(2) & (4):	Days of Maternity Leave		
	First Stage	IV	First Stage	IV LIML
	(1)	(2)	(3)	(4)
Gender Differentiated Personal Pronouns	7.09 (2.14)***		3.62 (2.11)*	
Men Have More Right to a Job		-2.34 (1.21)*		-3.86 (1.93)**
GDP per capita			-0.000005 (0.000002)**	-0.0017 (0.0013)
Government Share of GDP			0.002 (0.005)	3.89 (3.22)
Population			0.00 (0.00)	-0.00007 (0.00002)***
Women Share of Parliament			-0.0098 (0.0025)***	-1.71 (2.44)
Fertility			0.022 (0.045)	-6.95 (18.76)
R^2	0.17		0.56	
Instrument F -Statistic	10.98		2.94	
Number of obs.	55	55	55	55

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Standard errors are robust and clustered at the language level

attitudes on length of leave.

Table 3.8 presents the results of the instrumental variable regression, with standard errors that are robust and clustered at the language level.

Specification (1) presents the first stage regression, where the dependent variable is the percentage of people in a country who think that when jobs are scarce men should have more right to a job than women. We can see that there is a strong, statistically significant positive correlation between our linguistic measure and attitudes. Specifically, for every additional case of gender differentiated pronouns the share of people in the population who think that men have more right to a job than women increases by approximately 7.1 percent.

Specification (2) presents the second stage regression, where the dependent variable is the number of days of maternity leave, and we instrument for attitudes with our linguistic

measure. With this instrument we get a strong, statistically significant negative correlation between attitudes and the length of leave. For every additional percent of respondents who think that men should have more right to a job than women the length of maternity leave decreases by 2.3 days.

In specifications (3) and (4) we repeat the exercise in specification (1) and (2), only with all the controls. Note that in specification (3), the instrument is weak (F below 10). Since the Two Stage Least Squares has been found to have poor properties in such a settings, we use in specification (4) Limited Information Maximum Likelihood Estimation.⁸ We get again a strong, statistically significant negative correlation between attitudes, instrumented with our linguistic measure, and the length of leave.⁹

Comparing tables 3.8 and 3.2, one notices that the IV estimates are larger than the OLS estimates. This however is relatively common (see, for example, the return on schooling literature, reviewed in Card 2001), and can be explained by the fact that the use of IV may reduce the downward attenuation bias due to measurement error, or by the fact that in the set of countries which the IV estimates are identified off, that is countries where attitudes vary because of a change in our linguistic measure, the effect of attitudes on the length of leave is higher than in the set of all countries used for the OLS estimates.

The results in Table 3.8 support our claim that lower tolerance of gender based discrimination is associated with a longer mandated maternity leave.

3.4 Discussion

In this section we discuss some concerns that might arise with respect to our model and our results.

⁸For the literature on weak instruments see, among others, Buse (1992), Bound, Jaeger and Baker (1995), and Steiger and Stock (1997).

⁹One possible concern could be that other cultural variables that are affected by our linguistic measure are having an independent effect on the length of maternity leave. As a falsification test, we perform the same analysis for other World Values Survey variables that we found being correlated with our linguistic measure in the within country framework. We find that none of these variable is correlated with length of leave, whether with OLS or with IV. Results are available upon request.

Paternal Leave In our model we assumed that mandated leave is provided only to women and not to men. Since some countries mandate paternity leave as well as maternity leave, one can question the validity of our assumption.

It is important to note that our results do not hinge on the assumption that paternity leave is not mandated. What matters is that mandated paternity leave is shorter than mandated maternity leave, which makes hiring women rather than men relatively more costly. Thus, our assumption that mandated leave is provided only to women and not to men is simply a normalization.

Turning to the data, if what matters is the difference between the lengths of maternity leave and paternity leave, one can argue that our use of maternity leave as the dependent variable in tables 3.2 and 3.7 is incorrect. However, it turns out that accounting for paternity leave does not change our results. The reason is that most countries do not mandate a paternity leave, and those which do mandate only a very short one.

In table 3.9 we replicate tables 3.2 and 3.7, but instead of using the length of maternity leave as the dependent variable, we use the difference between the lengths of maternity leave and paternity leave, where data on paternity leave is taken from the Maternity Protection Database of the International Labor Organization (see appendix C.2 for source information).

One can see from table 3.9 that accounting for paternity leave does not affect our empirical results.

Multilingual Countries Our coding associates each country with the language most commonly spoken in it. This may pose a problem for some of the truly multilingual countries.

To address this issue we redefined, where possible, the number gender differentiated personal pronouns in these countries as a weighted average of the number of gender differentiated pronouns in the languages spoken in these countries, with weights given by the percentage of the population actually speaking each language. Using these weighted

Table 3.9: *Difference between Maternity and Paternity Leave*

Dependent Variable:	Difference between Days of Maternity Leave and Days of Paternity Leave			
	(1)	(2)	(3)	(4)
Gender Differentiated Personal Pronouns	−19.32 (5.61)***	−16.19 (4.21)***		
Men Have More Right to a Job			−1.34 (0.45)***	−0.80 (0.36)**
GDP per capita		0.0001 (0.0005)		0.0001 (0.0008)
Government Share of GDP		1.90 (1.07)*		3.55 (1.47)**
Population		−0.00007 (0.00001)***		−0.00005 (0.00001)***
Women Share of Parliament		1.82 (1.08)		1.43 (1.147)
Fertility		−8.94 (6.58)		−29.57 (9.30)***
R^2	0.14	0.27	0.13	0.24
Number of obs.	81	81	73	73

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Standard errors are robust, and clustered at the language level in (1) and (2)

averages does not change the results we get.¹⁰ For simplicity we choose to use the language most commonly spoken in our coding.

Arab Countries A concern that might arise is that our results are driven by countries where Arabic is spoken. As appendix C.3 notes, Arabic has relatively many cases of gender differentiated personal pronouns, and there are relatively many countries where Arabic is spoken. The concern is that if our results are driven by countries where Arabic is spoken they might be capturing another channel that is unique to these countries and that we do not control for, rather than attitudes towards gender based discrimination.

However, dropping the countries where Arabic is spoken from our sample does not

¹⁰The countries in the sample affected by this weighting are Belgium, Canada and Switzerland. If a language in a country is not coded it receives a zero weight. Note that for some multilingual countries where the different spoken languages have the same number of gender differentiated pronouns, such as Ukraine, where both Ukrainian and Russian have one case of gender differentiated pronouns, this procedure has no effect.

change our results. With these countries dropped there is still a strong negative correlation between the length of maternity leave and our linguistic measure, that is economically and statistically significant. Results are available upon request.

Direct Effect of Attitudes on Leave In the standard mandated benefit model, which we present in the beginning of section 3.2, women do not benefit from the introduction of mandated maternity leave, as the value of the mandated leave is smaller than their resulting decrease in wages. Furthermore, an increase in the length of mandated leave only harms women (since $c''(\mu) - v''(\mu) > 0$). Thus, in the standard model the more you care about women the shorter you want the mandated leave to be.

This result is important, since one could argue that our empirical results are not capturing the dynamics of our model, but rather a more simple effect. That is, one can argue that intolerance of gender based discrimination is simply a reflection of positive attitudes towards women, and that our empirical results simply show that countries with a more favorable attitudes towards women mandate a longer maternity leave. However, as just noted, in the standard mandated benefit model the more you care about women the shorter you want the mandated leave to be, which is inconsistent with this alternative explanation to our results.

3.5 Conclusion

Why do some countries mandate that employers provide a long maternity leave, while others mandate only a short one? This question seems particularly important as maternity leave is one of the key policies that supports continued labor force participation of women. We incorporate attitudes towards gender based discrimination into a standard mandated benefit model where employers provide maternity leave to women and not to men, showing that the less society is willing to tolerate gender based discrimination the longer the maternity leave it will provide.

Using the linguistic relativity principle we capture attitudes towards gender based discrimination with new data on the number of cases of gender differentiated personal

pronouns across languages. Using this measure we find, in a cross sectional analysis, results that are consistent with the prediction of our model.

References

Chapter 1

- Aghion, P., Alesina, A., and F. Trebbi (2004): "Endogenous Political Institutions," *Quarterly Journal of Economics*, 119, 565- 612.
- , and P. Bolton (1990): "Government Domestic Debt and the Risk of Default: a Political-Economic model of the Strategic Role of Debt," in R. Dornbusch and M. Draghi (eds.), *Public Debt Management: Theory and History*, Cambridge, UK: Cambridge University Press.
- Alesina, A., and T. Bayoumi (1996): "The Costs and Benefits of Fiscal Rules: Evidence from U.S. States," NBER Working Paper 5614.
- , and R. Perotti (1996): "Fiscal Discipline and the Budget Process," *American Economic Review Papers and Proceedings*, 86, 401–407.
- , and R. Perotti (1999): "Budget Deficits and Budget Institutions," in J. Poterba and J. Von Hagen (eds.), *Fiscal Institutions and Fiscal Performance*, Chicago, IL: University of Chicago Press.
- , and G. Tabellini (1990): "A Positive Theory of Fiscal Deficits and Government Debt," *Review of Economic Studies*, 57, 403–414.
- , Hausmann, R., Hommes, R., and E. Stein (1999): "Budget Institutions and Fiscal Performance in Latin America," *Journal of Development Economics*, 59, 253–273.
- Alt, J., and R. Lowry (1994): "Divided Government, Fiscal Institutions, and Budget Deficits: Evidence from the States," *American Political Science Review*, 88, 811–828.
- Angrist, J.D., and A.B. Krueger (1999): "Empirical Strategies in Labor Economics," in: O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. 3, Chapter 23, pp. 1277–1366, Elsevier.
- Auerbach, A. (2006): "Budget Windows, Sunsets, and Fiscal Control," *Journal of Public Economics*, 90, 87–100.
- Azzimonti, M., Battaglini, M., and S. Coate (2008): "Analyzing the Case for a Balanced Budget Amendment to the U.S. Constitution," mimeo.

- Balassone, F., Franco, D., and S. Zotteri (2004): "Fiscal Rules for Sub-national Governments: Lessons from EMU Countries," in G. Kopits (ed.), *Rules-Based Fiscal Policy in Emerging Markets: Background, Analysis and Prospects*, Palgrave Macmillan.
- Bandiera, O., Prat, A., and T. Valletti (2009): "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment," *American Economic Review*, 99, 1278–1308.
- Banks, J., and R.K. Sundaram, (1998): "Optimal Retention in Agency Problems," *Journal of Economic Theory*, 82, 293–323.
- Barro, R.J., (1973): "The Control of Politicians: An Economic Model," *Public Choice*, 14, 19–42.
- (1974): "Are Government Bonds Net Wealth?" *Journal of Political Economy*, 82, 1095–1117
- (1979): "On the Determination of the Public Debt," *Journal of Political Economy*, 87, 940–71.
- Bassetto, M., and T.J. Sargent (2006): "Politics and Efficiency of Separating Capital and Ordinary Government Budgets," *Quarterly Journal of Economics*, 121, 1167–1210.
- Battaglini, M., and S. Coate (2008): "A Dynamic Theory of Public Spending, Taxation, and Debt," *American Economic Review*, 98, 201–236.
- Bayoumi, T., and B. Eichengreen (1995): "Restraining Yourself: The Implications of Fiscal Rules for Economic Stabilization," *IMF Staff Papers*, 42, 32–48.
- Besley, T., and A. Case (1995): "Does Electoral Accountability Affect Economic Policy Choices? Evidence from Gubernatorial Term Limits," *Quarterly Journal of Economics*, 110, 769–798.
- (2007): *Principled Agents? The Political Economy of Good Government*, Oxford: Oxford University Press.
- , and M. Smart (2007): "Fiscal restraints and voter welfare," *Journal of Public Economics*, 91, 755–773.
- Bisin, A., A. Lizzeri, and L. Yariv (2011): "Government Policy with Time Inconsistent Voters," mimeo, New York University.
- Bohn, H., and R. Inman (1996): "Balanced-Budget Rules and Public Deficits: Evidence from the U.S. States," *Carnegie-Rochester Conference Series on Public Policy*, 45, 13–76.
- Bordignon, M., Nannicini, T., and G. Tabellini (2012): "Moderating Political Extremism: Single Round vs. Runoff Elections under Plurality Rule," mimeo, Bocconi University.
- Braun, M., and M. Tommasi (2004): "Subnational Fiscal Rules: A Game Theoretic Approach," in Kopits, G. (ed.), *Rules-Based Fiscal policy in Emerging Markets: Background, Analysis and Prospects*, Palgrave Macmillan.

- Calonico, S., M. D. Cattaneo, and R. Titiunik (2012): "Robust Nonparametric BiasCorrected Inference in the Regression Discontinuity Design", Working paper, University of Michigan .
- Casaburi, L., and U. Troiano (2012): "Ghost-House Busters: The Electoral Response to a Large Anti Tax Evasion Program," mimeo, Harvard University.
- Cellini, S.R., F. Ferreira, and J. Rothstein (2011): "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design," *Quarterly Journal of Economics*, 125, 215–261.
- Chetty, R., Looney, A., and K. Kroft (2009): "Salience and Taxation: Theory and Evidence," *American Economic Review*, 99, 1145–77.
- Clemens, J. (2012): "State Fiscal Adjustment During Times of Stress: Possible Causes of the Severity and Composition of Budget Cuts," mimeo.
- , and S. Miran (2012): "The Effects of State Budget Cuts on Employment and Income," *American Economic Journal: Economic Policy*, 4, 46–68.
- Debrun, X., L. Moulin, A. Turrini, J. Ayuso-i-Casals, and M.S. Kumar (2008): "Tied to the Mast? National Fiscal Rules in the European Union," *Economic Policy*, 54, 297–362.
- DellaVigna, S., and E. La Ferrara (2012): "Detecting Illegal Arms Trade," *American Economic Journal: Economic Policy*, forthcoming.
- Dickert-Conlin, S., and T. Elder (2010): "Suburban Legend: School Cutoff Dates and the Timing of Births," *Economics of Education Review*, 29, 826–841.
- Drazen, A. (2002): "Fiscal Rules From a Political Economy Perspective," paper prepared for the IMF–World Bank Conference *Rules-Based Fiscal Policy in Emerging Market Economies*, Oaxaca, Mexico, February 14–16.
- Eichengreen, B., and von Hagen, J. (1996): "Fiscal Policy and Monetary Union: is There a Trade-Off Between Federalism and Budgetary Restrictions?," NBER Working Paper 5517.
- Fatás, A., and I. Mihov (2006): "The Macroeconomic Effects of Fiscal Rules in the US States," *Journal of Public Economics*, 90, 101–117.
- Gagliarducci, S., and T. Nannicini (2012): "Do Better Paid Politicians Perform Better? Disentangling Incentives from Selection," *Journal of the European Economic Association*, forthcoming.
- Gavin, M., and R. Perotti (1997), "Fiscal policy in Latin America," *NBER Macroeconomics Annual* 12, 11–72.
- Givati, Y., and U. Troiano (2012): "Law, Economics and Culture: Theory and Evidence from Maternity Leave Laws," *Journal of Law and Economics*, forthcoming.

- Glaeser, E. (2012): "Urban Public Finance," in preparation for the *Handbook of Public Economics*.
- Hahn, J., P. Todd, and W. Van der Klaauw (2001): "Identification and Estimation of Treatment Effects with Regression Discontinuity Design," *Econometrica*, 69, 201–209.
- Hallerberg, M., and J. Von Hagen (1999): "Electoral Institutions, Cabinet Negotiations, and Budget Deficits in the EU," in J. Poterba and J. Von Hagen (eds.), *Fiscal Institutions and Fiscal Performance*, Chicago, IL: University of Chicago Press.
- Ichino, A., and G. Maggi (2000): "Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm," *Quarterly Journal of Economics*, 115, 1057–90.
- Imbens, G., and T. Lemieux (2008): "Regression Discontinuities Designs: A Guide to Practice," *Journal of Econometrics*, 142, 615–635.
- International Monetary Fund (2009): *Anchoring Expectations for Sustainable Public Finances*, Washington DC.
- Kaufmann, D., Kraay, A., and M. Mastruzzi (2010): "The Worldwide Governance Indicators: Methodology and Analytical Issues," World Bank Policy Research Working Paper 5430.
- Knight, B.G. (2000): "Supermajority Voting Requirements for Tax Increases: Evidence from the States," *Journal of Public Economics*, 76, 41–67.
- Kontopoulos, Y., and R. Perotti (1999): "Government Fragmentation and Fiscal Policy Outcomes: Evidence from OECD Countries," in J. Poterba and J. Von Hagen (eds.), *Fiscal Institutions and Fiscal Performance*, Chicago, IL: University of Chicago Press.
- Lemieux, T., and K. Milligan (2008): "Incentive Effects of Social Assistance: A Regression Discontinuity Approach," *Journal of Econometrics*, 142, 807–828.
- Lee, D.S., and T. Lemieux (2010): "Regression Discontinuities Designs in Economics," *Journal of Economic Literature*, 48, 281–355.
- List, J.A., and D.M. Sturm (2006): "How Elections Matter: Theory and Evidence from Environmental Policy," *Quarterly Journal of Economics*, 121, 1249–1281.
- Lizzeri, A. (1999): "Budget Deficits and Redistributive Politics," *Review of Economic Studies*, 66, 909–928.
- Lucas, R., and N. Stokey (1983): "Optimal Fiscal and Monetary Policy in an Economy without Capital," *Journal of Monetary Economics*, 12, 55–94.
- Maskin, E., and J. Tirole (2004): "The Politician and the Judge: Accountability in Government," *American Economic Review*, 94, 1034–1054.
- McCrary, J. (2008): "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 142, 698–714.

- Milesi-Ferretti, G.M. (2003): "Good, Bad or Ugly? On the Effects of Fiscal Rules on Creative Accounting," *Journal of Public Economics*, 88, 377–394.
- Nannicini, T., A. Stella, G. Tabellini, and U. Troiano (2012): "Social Capital and Political Accountability," *American Economic Journal: Economic Policy*, forthcoming.
- Olson, M. (1965): *The Logic of Collective Action*, Cambridge, MA: Harvard University Press.
- Persson, T., and G. Tabellini (2000): *Political Economics*, Cambridge, MA: MIT Press.
- , and L. Svensson (1989): "Exchange Rate Variability and Asset Trade," *Journal of Monetary Economics*, 23, 485–509.
- Pettersson-Lidbom, P. (2001): "An Empirical Investigation of the Strategic Use of Debt," *Journal of Political Economy*, 109, 570–584.
- (2012): "Does the Size of the Legislature Affect the Size of Government: Evidence from Two Natural Experiments," *Journal of Public Economics*, 96, 269–278.
- Poterba, J. (1994): "State Responses to Fiscal Crises: The Effects of Budgetary Institutions and Politics," *Journal of Political Economy*, 102, 799–821.
- (1996): "Budget Institutions and Fiscal Policy in the U.S. States," *American Economic Review Papers and Proceedings*, 86, 395–400.
- , and J. Von Hagen (eds) (1999): *Fiscal Institutions and Fiscal Performance*, Chicago, IL: University of Chicago Press.
- Rodden, J.A., G.S. Eskeland, and J. Litvack (eds) (2003): *Fiscal Decentralization and the Challenge of Hard Budget Constraints*, Cambridge, MA: MIT Press.
- Rogoff, K. (1990): "Equilibrium Political Budget Cycles," *American Economic Review*, 80, 21–36.
- Roubini, N., and J. Sachs (1989): "Political and Economic Determinants of Budget Deficits in the Industrial Democracies," *European Economic Review*, 33, 903–938.
- Song, Z., K. Storesletten, and F. Zilibotti (2012): "Rotten Parents and Disciplined Children: A Dynamic Politico-Economic Theory of Public Expenditure and Debt," *Econometrica*, forthcoming.
- Sutherland, D., R. Price, and I. Joumard (2005): "Fiscal Rules for Sub-Central Governments: Design and Impact," OECD Economics Department Working Paper 465.
- Tabellini, G., and A. Alesina (1990): "Voting on the Budget Deficit," *American Economic Review*, 80, 37–49.
- Ter-Minassian, T. (2007): "Fiscal Rules for Sub-national Governments: Can They Promote Fiscal Discipline?" *OECD Journal on Budgeting*, 6(3).
- Van der Klaauw, W. (2008): "Regression-Discontinuity Analysis: A Survey of Recent Developments in Economics," *Labour*, 22, 219–245.

- Von Hagen, J. (1991): "A Note on the Empirical Effectiveness of Formal Fiscal Restraints," *Journal of Public Economics*, 44, 199–210.
- Wyplosz, C. (2012), "Fiscal Rules: Theoretical Issues and Historical Experiences," NBER Working Paper 17884.
- Yared, P. (2010): "Politicians, Taxes and Debt," *The Review of Economic Studies*, 77, 806–40.

Chapter 2

- Alesina, Alberto (1988). "Credibility and Policy Convergence in a Two-Party System with Rational Voters." *American Economic Review*, 78, pp.796-806.
- , Reza Baqir and Caroline Hoxby (2004). "Political jurisdictions in heterogeneous communities." *The Journal of Political Economy*, 112(2):348-396.
- , Dorian Carloni, and Giampaolo Lecce (2011). "The Electoral Consequence of Large Fiscal Adjustments." Working paper.
- , William Easterly and Janina Matuszeski (2011). "Artificial States." *Journal of the European Economic Association*, 9(2): 246-277.
- and Enrico Spolaore (1997). "On the number and size of nations." *The Quarterly Journal of Economics*, 112(4):1027-1056.
- Alm, James, Betty Jackson, and Michael McKee (1992). "Estimating the Determinants of Taxpayer Compliance with Experimental Data." *Economic Development and Cultural Change*, March, 39(4): 107-114.
- Almunia, Miguel and David Lopez-Rodriguez (2012). "The Efficiency Costs of Tax Enforcement: Evidence from a Panel of Spanish Firms." University of California Berkeley, working paper.
- Andreoni, James, Brian Erard and Jonathan Feinstein (1998). "Tax Compliance." *Journal of Economic Literature*, 36(2):818-860.
- Aparicio, Gabriela (2012). "Monitoring and its Interaction with Punishment in Tax Enforcement: Evidence from a Regression Discontinuity Design." Georgetown University, working paper.
- Artavanis, Nikolaos T., Adait Morse, and Margarita Tsoutsoura (2012). "Tax Evasion Across Industries: Soft Credit Evidence from Greece." Chicago University Booth School of Business, working paper.
- Banks Jeffrey and Rangarajan Sundaram (1998). "Optimal Retention in Agency Problems." *Journal of Economic Theory*, 82(2). 293-323.
- Barca, Marcello (2008). "I primi cittadini della montagna commentano i dati apparsi sul sito dell'Agenzia del territorio sugli immobili non denunciati." *L'Informazione Provincia Montagna*, March 21st.

- Barro, Robert (1973). "The Control of Politicians: An Economic Model." *Public Choice*, 14, 1942.
- Bernardini, Daniele (2011). "Il Comune va a caccia delle case fantasma." *La Nazione* (2011), March 18th.
- Besley, Tim and Andrea Pratt (2006). "Handcuffing the Grabbing Hand?: Media Capture and Government Accountability." *American Economic Review*, 96(3). 720-736.
- , and Stephen Coate (1997). "An Economic Model of Representative Democracy." *The Quarterly Journal of Economics*, 112, pp.85-114.
- , Jose G. Montalvo, and Marta Reynal-Quenol (2012). "Do Educated Leaders Matter?" *Economic Journal*, forthcoming.
- Bird, Richard M. and Eric M. Zolt (2008). "Technology and Taxation in Developing Countries: From Hand to Mouse." *National Tax Journal*, Vol. LXI, No. 4, Part 2.
- Brender, Anton (2003). "The Effect of Fiscal Performance on Local Government Election Results in Israel: 1989-1998." *Journal of Public Economics*, 87 (9): 2187-2205.
- and Allan Drazen (2008). "How Do Budget Deficits and Economic Growth Affect Reelection Prospects? Evidence from a Large Panel of Countries." *American Economic Review*, 98(5): 2203-20.
- Bubb, Ryan (2008). "Blame It On The Rain? Voter Rationality and Exogenous Economic Shocks." New York University, working paper.
- Card, David (2001). "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica*, 69, pp. 1127-1160.
- Carpenter, Jeffrey, Samuel Bowles, Herbert Gintis, and Sung-Ha Hwang (2009). "Strong Reciprocity and Team Production: Theory and Evidence." *Journal of Economic Behavior and Organization*.
- Carrillo, Paul E., Dina Pomeranz, and Monica Singhal (2012). "Tax Me If You Can: Information Cross-Checks and Evasion Substitution." Working paper.
- Casari, Marco and Luigi Luini (2009). "Group cooperation under alternative punishment institutions: an experiment." *Journal of Economic Behavior and Organization*, 71(2). 273-282.
- Caselli, Francesco, Tom Cunningham, Massimo Morelli, and Inés Moreno-de-Barreda (2013). "The Incumbency Effects of Signalling." *Economica*, forthcoming.
- Cavallaro, Francesco (2011). "Sindaci Pronti alla Lotta all'Evasione." *Il Sole 24 Ore*. February 16th.
- Chetty, Raj, John Friedman and Emmanuel Saez (2011). "Using Neighborhood Effects to Uncover the Impacts of Tax Policy: The Effect of the EITC on Earnings." Working paper.

- Cole, Shawn, Andrew Healy, and Eric Werker (2012). "Do Voters Demand Responsive Governments? Evidence from Indian Disaster Relief." *Journal of Development Economics*, no. 97 (2012): 167-181.
- Comune di Capaccio Paestum (2010). "Comunicato Stampa, n. 134/10", July 29th.
- Corriere della Citta' (2012). "Crisi? Parla il sindaco De Fusco." March 28th.
- Dell'Oste, Cristiano and Gianni Trovati (2011). "Il Tesoro delle Case Fantasma." *Il Sole 24 Ore*, January 31st.
- Ferejohn, John (1986). "Incumbent performance and electoral control." *Public Choice*, 50: 5-26.
- Ferraz, Claudio, and Frederico Finan (2008) "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics*, 123 (2): 703-745.
- Fisman, Raymons and Shang-Jim Wei (2004). "Tax Rates and Tax Evasion: Evidence from Missing Imports in China." *Journal of Political Economy*, 112 (471-500).
- Gagliarducci, Stefano and Daniele Paserman (2012). "Gender Interactions within Hierarchies: Evidence form the Political Arena." *Review of Economic Studies*, 79(3). 1021-1052.
- . and Tommaso Nannicini (2013). "Do Better Paid Politicians Perform Better? Disentangling Incentives from Selection" , *Journal of the European Economic Association*, forthcoming
- Gazzetta del Mezzogiorno (2012). " Caccia aperta ai furbetti proprietari di case-fantasma. Il Comune vuole censire quelle abitazioni che sfuggono alle tasse." January 11th.
- Gordon, Roger H. and Wei Li. (2009). "Tax Structure in Developing Countries: Many Puzzles and a Possible Explanation." *Journal of Public Economics*, 93:7-8 (855-866).
- Grembi, Veronica, Tommaso Nannicini and Ugo Troiano (2013). "Policy Responses to Fiscal Restraints. A Difference-in-Discontinuities Design", Working paper
- Holmstrom, Berndt (1982). "Managerial incentive problems: A dynamic perspective." In *Essays in Economics and Management in Honor of Lars Wahlbeck*. Helsinki: Swedish School of Economics; reprinted in *Review of Economic Studies* 66 (1999): 169-182.
- Kleven Henrik J., Martin b. Knudsen, Claus T. Kreiner, Soren Pedersen , and Emmanuel Saez (2011). "Unwilling or Unable to Cheat? Evidence from a Randomized Tax Audit Experiment in Denmark." *Econometrica*, 79(3). 2011, 651-692.
- Kumler, Todd, Eric Verhoogen and Judith A. Frías (2011). "Enlisting Workers in Monitoring Firms: Payroll Tax Compliance in Mexico." Working paper.
- Lee, David. (2001). "The Electoral Advantage to Incumbency and Voters' Valuation of Politicians' Experience: a Regression Discontinuity Analysis of Elections to the U.S. House." NBER Working Paper #8441.

- Litschig, Stephan and Kevin Morrison (2012). "Government Spending and Re-election: Quasi-Experimental Evidence from Brazilian Municipalities." Barcelona GSE, working paper.
- Marino, M. Rosaria and Roberta Zizza (2008). "The personal income tax evasion in Italy." *Bank of Italy*, working paper.
- Mian Atif and Amir Sufi (2012). "The Effects of Fiscal Stimulus: Evidence from the 2009 Cash for Clunkers Program." *Quarterly Journal of Economics*, forthcoming.
- Miguel, Edward A., Marco Manacorda and Andrea Vigorito (2011). "Government Transfers and Political Support." *American Economic Journal: Applied Economics*, 3(3). 128.
- Nannicini, Tommaso, Andrea Stella, Guido Tabellini, and Ugo Troiano (2012). "Social Capital and Political Accountability." *American Economic Journal: Economic Policy*, forthcoming.
- Nordhaus, William (1975). "The political business cycle." *Review of Economic Studies*, 42:169-190.
- Olson, Mancur (1965). "The Logic of Collective Action: Public Goods and the Theory of Groups." Harvard University Press.
- Ouss Aurelie and Alex Peysakhovich "When Punishment Doesn't Pay: The 'Cold Glow' Heuristic and Decisions to Punish." Harvard University, working paper.
- Pande, Rohini (2011). "Can Informed Voters Enforce Better Governance? Experiments in Low-Income Democracies." *Annual Review of Economics*, Vol. 3: 215-237.
- Pomeranz Dina (2012). "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax." Harvard Business School, working paper.
- Putnam, Robert (2001). "Social Capital: Measurement and Consequences." *ISUMA - Canadian Journal of Policy Research*, 2, 1, 41-51.
- Quattrone, George A. and Amos Tversky. (1988). "Contrasting Rational and Psychological Analyses of Political Choice." *The American Political Science Review*, Vol. 82, No. 3., pp. 719-736.
- Rogoff, Kenneth, and Anne Sibert (1988). "Elections and macroeconomic policy cycles." *Review of Economic Studies*, 55:1-16.
- Rothstein, Bo (2000). "Trust, Social Dilemmas and Collective Memories." *Journal of Theoretical Politics*, vol. 12, 477-501.
- Saez, Emmanuel (2010). "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, 2, pp. 180-212.
- Saiz, Albert (2010). "The Geographic Determinants of Housing Supply." *The Quarterly Journal of Economics*, vol. 125(3). pages 1253-1296.

- Slemrod, Joel (2003). "Trust in Public Finance." In S. Cnossen and H. Werner (eds.) *Public Finance and Public Policy in the New Century*, 49–88. Cambridge, MA: MIT Press.
- (2007). "Cheating Ourselves: The Economics of Tax Evasion." *Journal of Economic Perspectives*, 21(1): 25-48.
- , Marsha Blumenthal, and Charles Christian (2001). "Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota." *Journal of Public Economics* 79(3): 455- 483.
- and Shlomo Yitzhaki (2002). "Tax avoidance, evasion, and administration." Handbook of Public Economics, in: A. J. Auerbach and M. Feldstein (ed.). *Handbook of Public Economics*, edition 1, volume 3, chapter 22, pages 1423-1470.
- Spicer, Michael W., and Lee A. Becker (1980). "Fiscal Inequity and Tax Evasion: An Experimental Approach." *National Tax Journal*, June, 33(2):171-175.
- Torgler, Benno (2003). "Tax Morale, Rule-Governed Behaviour, and Trust." *Constitutional Political Economy*, June, 14(2): 119-140.
- (2007). *Tax Morale and Tax Compliance: A Theoretical and Empirical Analysis*, Cheltenham U.K.: Edward Elgar.
- Tullock, Gordon (1959). "Some problems of majority voting." *Journal of Political Economy*, 67: 571-579.
- Wolfers, Justin (2009). "Are Voters Rational? Evidence from Gubernatorial Elections." University of Pennsylvania, working paper.

Chapter 3

- Aghion, Philippe, Alberto Alesina and Francesco Trebbi (2004). "Endogenous Political Institutions." *Quarterly Journal of Economics* 119: 565-612.
- , Yann Algan, Pierre Cahuc and Andrei Shleifer (2010). "Regulation and Distrust." *Quarterly Journal of Economics* 125: 1015-1049.
- Alesina, Alberto and Paola Giuliano (2010). "The Power of the Family." *Journal of Economic Growth* 15: 93-125.
- Bisin, Alberto and Thierry Verdier (2000). "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits." *Quarterly Journal of Economics* 115: 955-988.
- and Thierry Verdier (2001). "The Economics of Cultural Transmission and the Evolution of Preferences." *Journal of Economic Theory* 97: 298-319.
- Boroditsky, Lera (2003). "Linguistic Relativity," in *Encyclopedia of Cognitive Science*, ed. Lynn Nade (MacMillan Press), 917-921.
- (2010). "Lost in Translation." *Wall Street Journal*, July 23.

- (2011). "How Language Shapes Thought." *Scientific American*, February, 63-65.
- , Lauren A. Schmidt and Webb Phillips (2003). "Sex, Syntax, and Semantics," in *Language in Mind: Advances in the study of Language and Cognition*, ed. Dedre Gentner and Susan Goldin-Meadow (MIT Press), 61-79.
- Bound, John, David A. Jaeger and Regina M. Baker (1995). "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables Is Weak." *Journal of the American Statistical Association* 90: 443-450.
- Buse, A. (1992). "The Bias of Instrumental Variable Estimators." *Econometrica* 60: 173-180.
- Card, David (2001). "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69: 1127-1160.
- Deutscher, Guy (2010a). "Does Your Language Shape How You Think?" *New York Times Sunday Magazine*, August 29, MM42.
- (2010b). *Through The Language Glass: Why the World Looks Different in Other Languages* (Metropolitan Books).
- Diamond, Peter A. (1975). "A Many-Person Ramsey Tax Rule." *Journal of Public Economics* 4: 335-342.
- Dowd, Nancy E. (1989). "Envisioning Work and Family: A Critical Perspective on International Models." *Harvard Journal on Legislation* 26: 311-348.
- Fukuyama, Francis (1995). *Trust: The Social Virtues and the Creation of Prosperity* (Free Press).
- Givati, Yehonatan (2011). "The Comparative Law and Economics of Plea Bargaining: Theory and Evidence." Harvard Law School Olin Discussion Paper No. 39.
- Grill, Arielle H. (1996) "The Myth of Unpaid Family Leave: Can the United States Implement a Paid Leave Policy Based on the Swedish Model?" *Comparative Labor Law Journal* 17: 373-397.
- Guiso, Luigi, Paola Sapienza and Luigi Zingales (2006). "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives* 20: 23-48.
- , Paola Sapienza and Luigi Zingales (2008). "Social Capital as Good Culture." *Journal of the European Economic Association* 6: 295-320.
- , Paola Sapienza and Luigi Zingales (2009). "Cultural Biases in Economic Exchange?" *Quarterly Journal of Economics* 124: 1095-1131.
- Guiora, Alexander Z., Benjamin Beit-Hallahmi, Risto Fried and Cecelia Yoder (1982). "Language Environment and Gender Identity Attainment." *Language Learning* 32: 289-304.

- Gumperz, John J. and Stephen C. Levinson, ed. (1996). *Rethinking Linguistic Relativity* (Cambridge University Press)
- Issacharoff, Samuel and Elyse Rosenblum (1994). "Women and the Workplace: Accommodating the Demands of Pregnancy." *Columbia Law Review* 94: 2154-2221.
- Jolls, Christine (2000). "Accommodation Mandates." *Stanford Law Review* 53: 223-306.
- (2006). "Law and The Labor Market." *Annual Review of Law and Social Science* 2: 359-385.
- Hill, Jane H. and Bruce Mannheim (1992). "Language and World View." *Annual Review of Anthropology* 21: 381-406.
- Lalive, Rafael and Joseph Zweimüller (2009). "How Does Parental Leave Affect Fertility and Return to Work? Evidence from Two Natural Experiments." *Quarterly Journal of Economics* 124: 1363-1402.
- Levinson, Stephen C. (2003). *Space in Language and Cognition: Explorations in Cognitive Diversity* (Cambridge University Press).
- Levmore, Saul (2007). "Parental Leave and American Exceptionalism." *Case Western Reserve Law Review* 58: 203-222.
- Licht, Amir N., Chanan Goldschmidt, and Shalom H. Schwarz (2006). "Culture Rules: The Foundations of the Rule of Law and Other Norms of Governance," Mimeo, UC Berkely School of Law.
- Lucy, John A. 1997. "Linguistic Relativity." *Annual Review of Anthropology* 26: 291-312.
- and Suzanne Gaskins (2001). "Grammatical Categories and The Development of Classification Preferences: A Comparative Approach," in *Language Acquisition and Conceptual Development*, ed. Melissa Bowerman and Stephen C. Levinson (Cambridge University Press), 257-283.
- Nannicini, Tommaso, Andrea Stella, Guido Tabellini and Ugo Troiano (2010). "Social Capital and Political Accountability." CEPR Discussion Paper 7782.
- Pelletier, Annie (2006). "The Family Medical Leave Act of 1993—Why does parental leave in the United States fall so far behind Europe?" *Gonzaga Law Review* 42: 547-575.
- Putnam, Robert D. (1993). *Making Democracy Work: Civic Traditions in Modern Italy* (Princeton University Press).
- (2000). *Bowling Alone: The Collapse and Revival of American Community* (Simon and Schuster).
- Rasnic, Carol D. (1994) "United States' 1993 Family and Medical Leave Act: How Does It Compare with Work Leave Laws in European Countries." *Connecticut Journal of International Law* 10: 105-151.
- Ruhm, Christopher J. (1998). "The Economic Consequences of Parental Leave Mandates: Lessons from Europe." *Quarterly Journal of Economics* 113: 285-317.

- (2000). "Parental Leave and Child Health." *Journal of Health Economics* 19: 931-960.
- Sapir, Edward (1929). "The Status of Linguistics As a Science." *Language* 5: 207-214.
- Schuchmann, Mona L. (1995). "The Family and Medical Leave Act of 1993: A Comparative Analysis with Germany." *Journal of Corporation Law* 20: 331-361.
- Spamann, Holger (2009). "Large-Sample, Quantitative Research Designs for Comparative Law?" *American Journal of Comparative Law* 57: 797-810.
- Staiger, Douglas and James H. Stock (1997). "Instrumental Variables Regression With Weak Instruments." *Econometrica* 65: 557-586.
- Suk, Julie C. (2010). "Are Gender Stereotypes Bad for Women? Rethinking Antidiscrimination Law and Work-Family Conflict." *Columbia Law Review* 110: 1-69
- Summers, Lawrence H. (1989). "Some Simple Economics of Mandated Benefits." *American Economic Review* 79:177-183.
- Tabellini, Guido (2008). "Institutions and Culture." *Journal of the European Economic Association* 6: 255-294.
- Whorf, Benjamin L. (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, ed. John B. Carroll (MIT Press).
- Winawer, Jonathan, Nathan Witthoft, Michael C. Frank, Lisa Wu, Alex R. Wade, and Lera Boroditsky (2007). "Russian Blues Reveal Effects of Language on Color Discrimination." *Proceedings of the National Academy of Sciences* 104(19): 7780-7785.

Appendix A

Appendix to Chapter 1

A.1 Supplementary Tables and Figures

This Appendix provides additional information and robustness checks, which are also discussed in the paper. In particular, we describe the characteristics and sources of the variables we use (Table A1), and we present further robustness checks:

- diff-in-disc estimates with covariates (Table A2);
- balance tests of time-invariant municipal characteristics (Table A3);
- diff-in-disc estimates on potentially endogenous variables (Table A4);
- falsification tests for the heterogeneity analysis (Table A5, Table A6, and Table A7);
- test of the continuity of the density at 5,000 in the 1991 Census, in the 2001 Census, and with respect to the difference between the two Censuses (Figure A1);
- placebo tests based on permutation methods (Figure A2 and Figure A3).

Table A.1: *Variables' description and sources*

Variable	Definition and measure	Available from-to	Source
<i>Deficit</i>	Expenditure minus revenues Per-resident; 2009 Euros	1997-2004	IMI
<i>Fiscal gap</i>	Expenditure minus revenues (net of central transfers and debt service) Per-resident; 2009 Euros	1998-2004	IMI
<i>Current outlays</i>	Total current expenditure Per-resident; 2009 Euros	1998-2004	IMI
<i>Capital outlays</i>	Total capital expenditure Per-resident; 2009 Euros	1998-2004	IMI
<i>Debt service</i>	Interest payments on outstanding debt Per-resident; 2009 Euros	1998-2004	IMI
<i>Taxes</i>	Total tax revenues Per-resident; 2009 Euros	1997-2004	IMI
<i>Fees & tariffs</i>	Total revenues from fees and tariffs Per-resident; 2009 Euros	1997-2004	IMI
<i>Central transfers</i>	Total transfers by the central state Per-resident; 2009 Euros	1997-2004	IMI
<i>Other revenues</i>	Residual category Per-resident; 2009 Euros	1997-2004	IMI
<i>Real estate tax rate</i>	The tax rate on real estate From 0.004 to 0.007 of the home value	1997-2004	IFEL-ANCI
<i>Income tax surcharge</i>	Municipal income tax surcharge Up to 0.6% of the taxable income	1999-2004	ME-DF

Notes: IMI stands for Italian Ministry of the Interior; IFEL-ANCI stands for Institute for the Local Finance and Economy of the National Italian Association of Municipalities; ME-DF stands for Italian Ministry of the Economy, Department of Finance.

Table A1 (contd.): Variables' description and sources

Variable	Definition and Measure	Available from-to	Source
<i>Census population</i>	Census population of the municipality	1991 and 2001	ISTAT
<i>Young cohorts</i>	Ratio of residents aged 0–14 over resident population Fraction at municipality level	1998-2004	ISTAT
<i>Speed of public good</i>	Paid over committed current expenditures Fraction at municipality level	1999-2004	IMI
<i>Area size</i>	Municipal area size In km ²	1999-2004	IMI
<i>Sea level</i>	Municipal sea level In meters	1999-2004	IMI
<i>Taxable income</i>	Municipal taxable income mean Per-resident; 2009 Euros	1999-2004	ME-DF
<i>Female Mayor</i>	Equal to 1 if the mayor in office is a woman Dummy variable	1999-2004	IMI
<i>Mayor's age</i>	Age of the mayor Number of years	1999-2004	IMI
<i>Mayor's schooling</i>	Years of choosing of the mayor in office Number of years	1999-2004	IMI
<i>Mayor's tenure</i>	Experience of the mayor in office Number of mandates	1999-2004	IMI
<i>Term limit</i>	Equal to 1 if the mayor in office faces term limit Dummy variable	1999-2004	IMI

Notes: IMI stands for Italian Ministry of the Interior; IFEL-ANCI stands for Institute for the Local Finance and Economy of the National Italian Association of Municipalities; ME-DF stands for Italian Ministry of the Economy, Department of Finance.

Table A.2: *The effect of relaxing fiscal rules, estimates with covariates*

	LLR $h = 500$	LLR $h = 750$	Spline poly 3 rd	Spline poly 4 th
A. Fiscal discipline				
Deficit	16.085** (8.082)	19.265*** (6.208)	21.094** (9.283)	23.396* (12.579)
Fiscal gap	70.704*** (26.436)	56.734*** (20.049)	96.748*** (32.595)	118.962*** (41.570)
B. Expenditures				
Current outlays	-60.558 (49.861)	-13.958 (32.864)	-44.755 (50.871)	-68.705 (72.117)
Capital outlays	54.450 (76.831)	44.429 (59.520)	102.644 (101.934)	224.342* (136.181)
Debt service	-1.962 (6.107)	-1.077 (3.833)	-1.962 (6.145)	-2.298 (9.151)
C. Revenues				
Taxes	-56.638** (23.920)	-40.683** (17.221)	-56.863** (24.024)	-93.975*** (30.751)
Fees & tariffs	-6.989 (9.676)	-2.700 (7.028)	-2.992 (10.049)	-10.241 (13.341)
Central transfers	52.658** (23.251)	36.392** (17.730)	73.692** (29.036)	93.268** (36.275)
Other revenues	-13.186 (92.763)	17.120 (67.138)	20.996 (112.937)	140.891 (155.552)
D. Tax instruments				
Real estate tax rate	-0.043* (0.024)	-0.030* (0.018)	-0.058** (0.026)	-0.060* (0.033)
Income tax surcharge	-0.027 (0.037)	-0.053* (0.029)	-0.051 (0.042)	-0.103** (0.052)
Obs.	2,080	3,068	6,300	6,300

Notes. Municipalities between 3,500 and 7,000 inhabitants; budget years between 1999 and 2004. Diff-in-disc estimates of the impact of relaxing fiscal rules on policy outcomes and tax instruments below 5,000 after 2001. Specifications augmented by controlling for covariates: yearly dummies, macro areas dummies (i.e. North West, North East, South), area size (in km²), and sea level (in meters). Estimation methods: Local Linear Regression (LLR) with bandwidth $h = 500$ or $h = 750$, as in equation (1.5); spline polynomial approximation with 3rd-order or 4th-order polynomial, as in equation (1.6). All policy outcomes are per capita and in 2009 Euros. Tax instruments are in percentage points. Robust standard errors clustered at the municipality level are in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table A.3: *Balance tests of time-invariant characteristics*

	LLR $h = 500$	LLR $h = 750$	Spline poly 3 rd	Spline poly 4 th
North-West	0.012 (0.109)	0.057 (0.088)	0.043 (0.122)	0.167 (0.152)
North-East	0.072 (0.088)	0.039 (0.071)	0.042 (0.095)	-0.039 (0.115)
Center	-0.043 (0.096)	-0.032 (0.079)	-0.071 (0.106)	-0.181 (0.133)
South	-0.041 (0.075)	-0.064 (0.064)	-0.013 (0.088)	0.053 (0.110)
Area size	-2.791 (9.045)	-0.519 (7.377)	0.043 (10.199)	-7.710 (11.734)
Sea level	-10.800 (44.481)	-9.053 (34.398)	-17.736 (46.537)	9.258 (59.757)
Obs.	2,080	3,068	6,300	6,300

Notes. Municipalities between 3,500 and 7,000 inhabitants. Diff-in-disc estimates with changing population levels (1991 Census before 2001 and 2001 Census after 2001). All time-invariant characteristics are dummies except area size (in km²) and sea level (in meters). Estimation methods: Local Linear Regression (LLR) with bandwidth $h = 500$ or $h = 750$; spline polynomial approximation with 3rd-order or 4th-order polynomial. Robust standard errors clustered at the municipality level are in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table A.4: *Balance tests of potentially endogenous characteristics*

	LLR $h = 500$	LLR $h = 750$	Spline poly 3 rd	Spline poly 4 th
Taxable income	211.702 (638.019)	-187.577 (491.318)	-230.188 (679.745)	-19.672 (857.344)
Female mayor	-0.078 (0.072)	-0.083 (0.064)	-0.112 (0.079)	-0.068 (0.095)
Mayor's age	0.457 (2.846)	-1.361 (2.166)	-0.770 (3.114)	-0.907 (3.930)
Mayor's schooling	0.646 (0.838)	0.698 (0.645)	0.818 (0.929)	1.192 (1.155)
Mayor's tenure	0.551 (1.378)	-0.081 (1.110)	0.444 (1.497)	1.287 (1.848)
No. of parties	-0.394 (0.440)	-0.200 (0.346)	-0.498 (0.485)	-0.958 (0.621)
Term limit	-0.097 (0.136)	-0.075 (0.107)	-0.136 (0.150)	-0.123 (0.184)
Young cohorts	-0.085 (0.141)	-0.083 (0.110)	-0.065 (0.154)	-0.083 (0.198)
Efficiency	0.052 (0.152)	0.044 (0.116)	-0.019 (0.165)	0.035 (0.209)
Obs.	2,080	3,068	6,300	6,300

Notes. Municipalities between 3,500 and 7,000 inhabitants. Baseline diff-in-disc estimates. Taxable income at the municipal level is per capita and in 2009 Euros; mayor's age, schooling, and tenure are expressed in years; female mayor and term limit are dummies; number of parties refer to political parties seating in the city council. See the Appendix Table A1 for a precise description of all these variables. Estimation methods: Local Linear Regression (LLR) with bandwidth $h = 500$ or $h = 750$; spline polynomial approximation with 3rd-order or 4th-order polynomial. Robust standard errors clustered at the municipality level are in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table A.5: *The political economy of deficit bias, part I – Falsification test*

	LLR $h = 500$	LLR $h = 750$	Spline poly 3 rd	Spline poly 4 th
With two parties or less:				
Deficit	4.342 (8.885)	-11.267 (10.361)	-2.061 (11.445)	2.368 (12.799)
Obs.	680	988	2,217	2,217
With more than two parties:				
Deficit	-2.905 (11.359)	-0.567 (10.595)	-1.459 (13.363)	-9.829 (15.094)
Obs.	580	864	1,963	1,963
<i>Difference between the two subsamples</i>	-7.247	10.700	0.602	-12.197
<i>Wald test p-value without covariates</i>	0.185	0.393	0.168	0.144
With binding term limit:				
Deficit	-4.222 (14.889)	0.764 (13.381)	2.291 (17.202)	-24.228 (22.308)
Obs.	388	550	1,213	1,213
Without binding term limit:				
Deficit	-2.862 (11.869)	-9.567 (10.969)	-4.463 (13.794)	4.983 (16.837)
Obs.	872	1,302	2,967	2,967
<i>Difference between the two subsamples</i>	1.360	-10.331	-6.754	29.211
<i>Wald test p-value without covariates</i>	0.945	0.559	0.762	0.303

Notes. Municipalities between 3,500 and 7,000 inhabitants; budget years between 1997 and 2001. Diff-in-disc estimates of the impact of the false relaxation of fiscal discipline below 5,000 after 1999 in different subsamples (that is, above vs. below median number of parties; binding vs. non-binding term limit). Estimation methods: Local Linear Regression (LLR) with bandwidth $h = 500$ or $h = 750$, as in equation (1.5); spline polynomial approximation with 3rd-order or 4th-order polynomial, as in equation (1.6). The *Wald test p-value without covariates* evaluates whether the estimates are statistically different in the two subsamples. The *Wald test p-value with covariates* is not available because of data limitations before 1999. All variables are per capita and in 2009 Euros. Robust standard errors clustered at the municipality level are in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table A.6: *The political economy of deficit bias, part II – Falsification test*

	LLR $h = 500$	LLR $h = 750$	Spline poly 3 rd	Spline poly 4 th
	Young cohorts above median:			
Deficit	4.124 (9.125)	-6.322 (9.933)	0.741 (11.431)	2.273 (12.944)
Obs.	805	1,160	2,599	2,599
	Young cohorts below median:			
Deficit	2.581 (15.266)	-0.816 (12.143)	3.824 (16.911)	-1.688 (21.845)
Obs.	455	692	1,581	1,581
<i>Difference between the two subsamples</i>	-1.543	5.506	3.083	-3.961
<i>Wald test p-value without covariates</i>	0.933	0.733	0.885	0.882

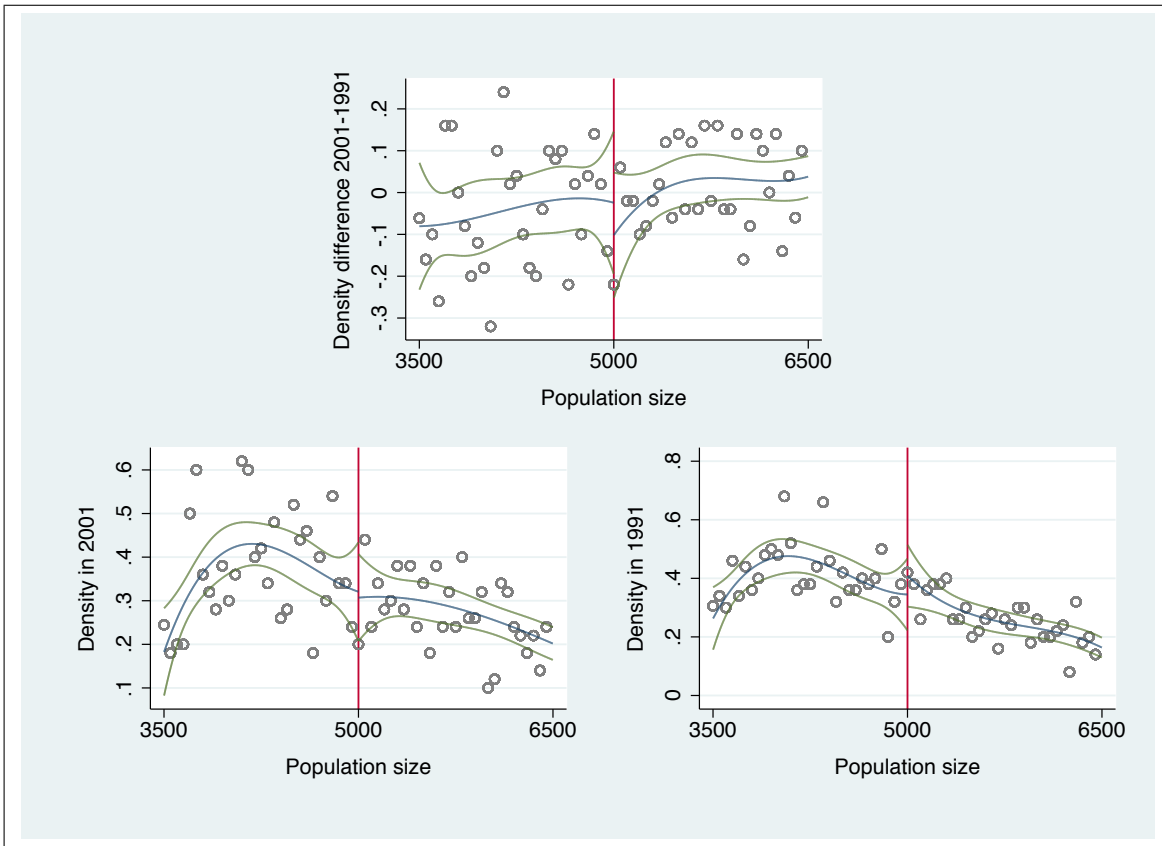
Notes. Municipalities between 3,500 and 7,000 inhabitants; budget years between 1997 and 2001. Diff-in-disc estimates of the impact of the false relaxation of fiscal discipline below 5,000 after 1999 in different subsamples (that is, above vs. below median percentage of young cohorts). Estimation methods: Local Linear Regression (LLR) with bandwidth $h = 500$ or $h = 750$, as in equation (1.5); spline polynomial approximation with 3rd-order or 4th-order polynomial, as in equation (1.6). The *Wald test p-value without covariates* evaluates whether the estimates are statistically different in the two subsamples. The *Wald test p-value with covariates* is not available because of data limitations before 1999. All variables are per capita and in 2009 Euros. Robust standard errors clustered at the municipality level are in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table A.7: *Fiscal restraints and budget management – Falsification test*

	LLR $h = 500$	LLR $h = 750$	Spline poly 3 rd	Spline poly 4 th
Speed of public good provision above median:				
Deficit	2.989 (9.812)	-2.653 (8.166)	6.135 (10.751)	0.430 (13.740)
Obs.	652	977	2,283	2,283
Speed of public good provision below median:				
Deficit	0.842 (11.906)	-10.780 (14.003)	-9.531 (15.465)	-3.501 (15.622)
Obs.	608	875	1,897	1,897
<i>Difference between the two subsamples</i>	-2.147	-8.127	-15.666	-3.931
<i>Wald test p-value without covariates</i>	0.883	0.607	0.381	0.840

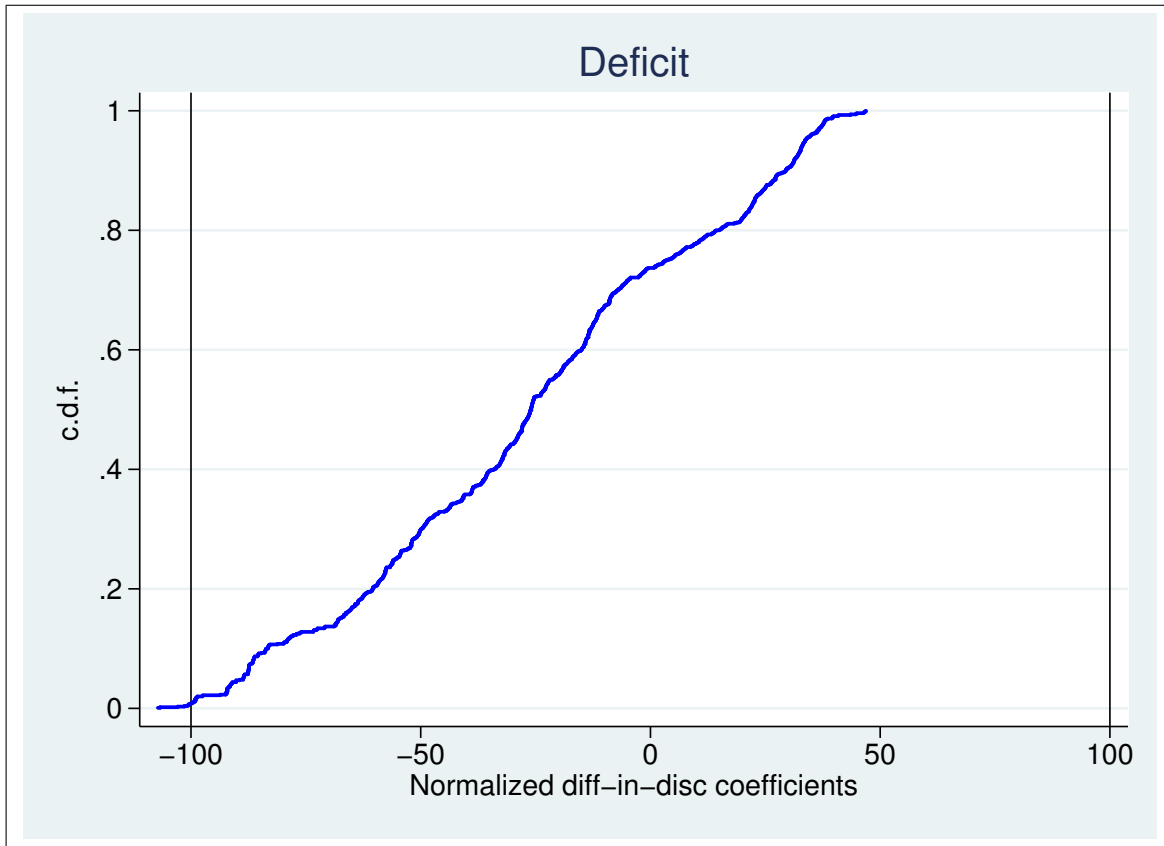
Notes. Municipalities between 3,500 and 7,000 inhabitants; between 1997 and 2001. Diff-in-disc estimates of the impact of the false relaxation of fiscal discipline below 5,000 after 1999 in different subsamples (that is, above vs. below median speed of public good provision). Estimation methods: Local Linear Regression (LLR) with bandwidth $h = 500$ or $h = 750$, as in equation (1.5); spline polynomial approximation with 3rd-order or 4th-order polynomial, as in equation (1.6). The *Wald test p-value without covariates* evaluates whether the estimates are statistically different in the two subsamples. The *Wald test p-value with covariates* is not available because of data limitations before 1999. All variables are per capita and in 2009 Euros. Robust standard errors clustered at the municipality level are in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Figure A1: Density tests



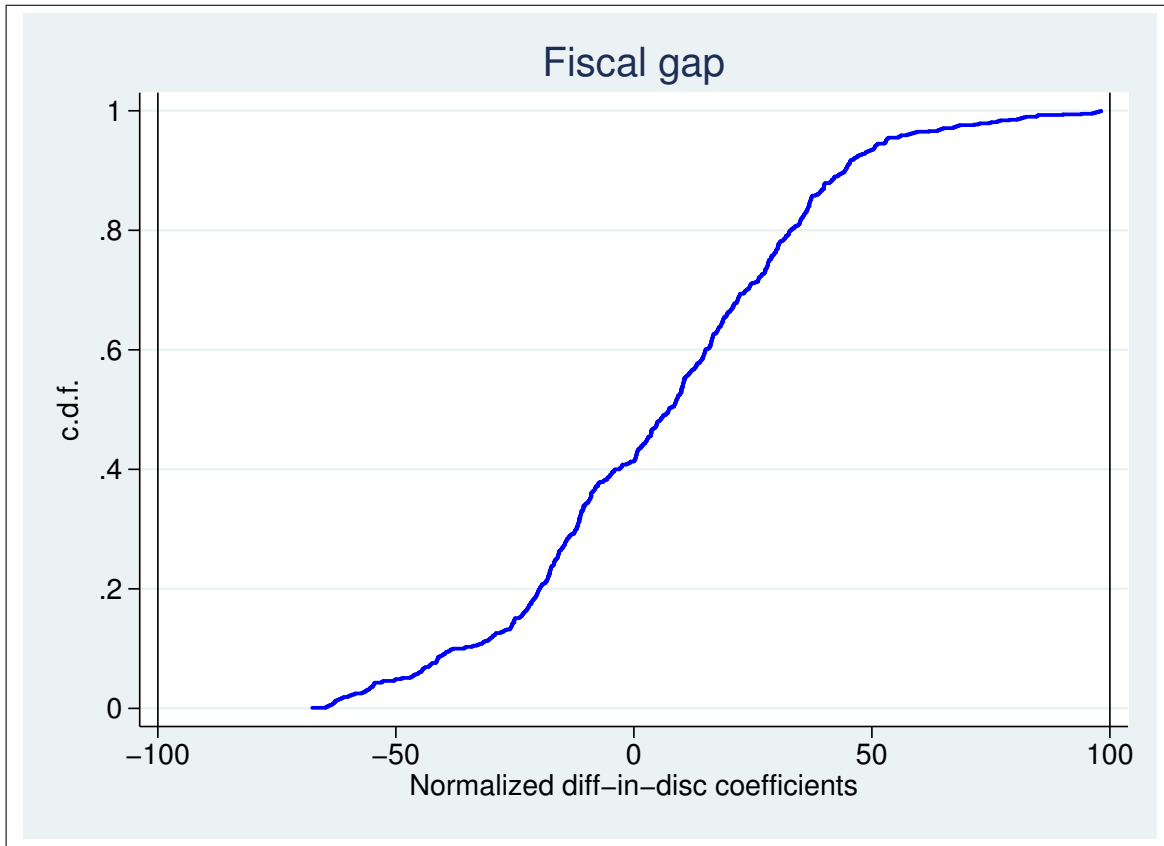
Notes. Test of the continuity at 5,000 of: (i) the difference between the density in the 2001 Census and in the 1991 Census (top graph); (ii) the density in the 2001 Census (bottom left graph); and (iii) the density in the 1991 Census (bottom right graph). The central line is a spline 3^{rd} -order polynomial fit in population size; the lateral lines represent the 95% confidence interval. Scatter points are averaged over intervals of 50 inhabitants.

Figure A2: Placebo tests for deficit



Notes. Placebo tests based on permutation methods for deficit. The figure reports the empirical c.d.f. of the normalized point estimates from a set of diff-in-disc estimations at 500 false thresholds below and 500 false thresholds above the true threshold at 5,000 (namely, any point from 4,900 to 4,400 and any point from 5,100 to 5,600). Estimation method: spline polynomial approximation with 3rd-order polynomial. The vertical lines indicate our benchmark estimate for deficit from Table 1.5 (i.e., true coefficient normalized to 100) and its negative value.

Figure A3: Placebo tests for fiscal gap



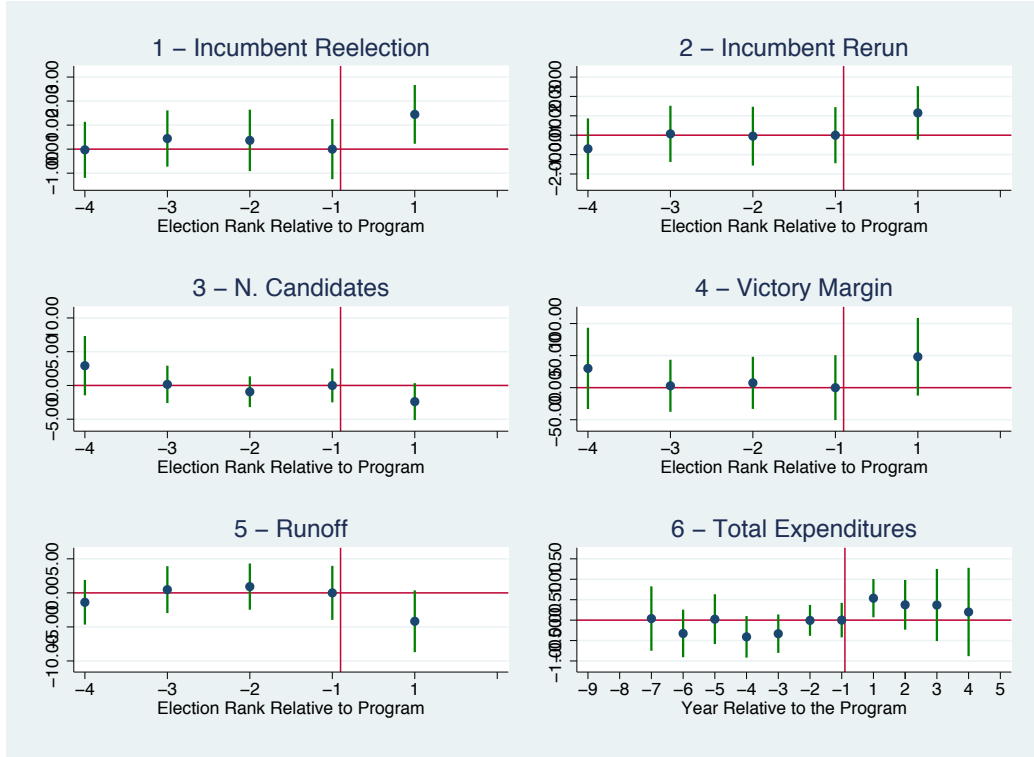
Notes. Placebo tests based on permutation methods for fiscal gap. The figure reports the empirical c.d.f. of the normalized point estimates from a set of diff-in-disc estimations at 500 false thresholds below and 500 false thresholds above the true threshold at 5,000 (namely, any point from 4,900 to 4,400 and any point from 5,100 to 5,600). Estimation method: spline polynomial approximation with 3rd-order polynomial. The vertical lines indicate the normalized benchmark estimate for fiscal gap from Table 1.5 (i.e., true coefficient normalized to 100) and its negative value.

Appendix B

Appendix to Chapter 2

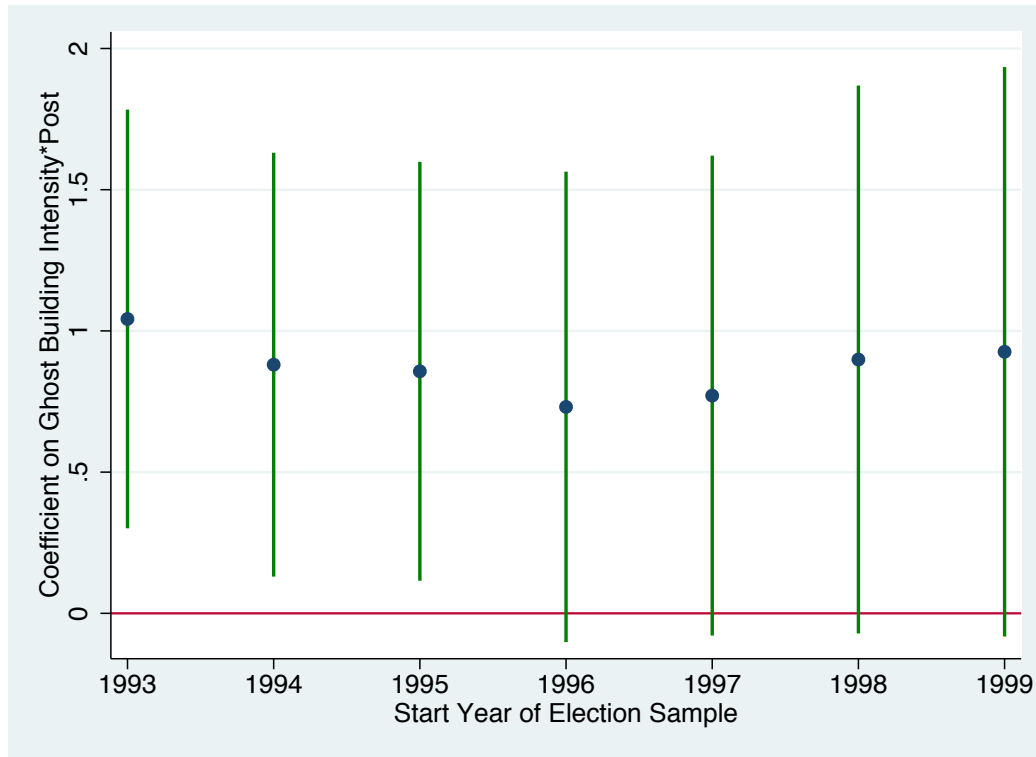
B.1 Appendix Figures

Figure B.1: *Ghost Building Intensity Coefficient by Election Pre/Post Program*



Notes: **Graphs 1 to 5** report the coefficients on the ghost building intensity *for each election* before and after the beginning of the Ghost Buildings program. The modal number of years between elections is five years between 1993 and 2001, and four afterwards. On the x-axis, elections are ranked based on their occurrence relative to the program. **Graph 6** reports the coefficients on the ghost building intensity *for each calendar year* before and after the beginning of the Ghost Buildings program. The dependent variable is in natural logarithm. On the x-axis, years are ranked based on their occurrence relative to the program. We drop the year of program inception due to its ambiguous treatment status. In all the graphs, the regression includes town and year fixed effects. We report the point estimate and the 95% confidence interval. The last election/year before the program (“-1”) is the omitted category. The coefficient on ghost building intensity for this group is normalized to zero. Confidence interval width for this election/year is obtained as the mean of the confidence interval width in election/year -2 and election/year +1.

Figure B.2: *Robustness to changes in election sample time span*



Notes: The figure presents robustness of the results on incumbent reelection to changes in sample years. The y-axis shows the coefficient (and 95% confidence intervals) on the interaction between ghost building intensity and the post-program indicator as estimated in our baseline specification (Table 2.4). The x-axis is the start year of the alternative election samples we use (the final year is 2011 for all the samples). The first sample, 1993-2011, is the baseline sample. In 1993, an electoral law reform introduced mayor individual ballot election.

B.2 Appendix Tables

Table B.1: *Variables' description and sources*

Variable	Definition and measure	Sample	Source
<i>Ghost Building Intensity</i>	Ghost Building Intensity Ratio between the number of land registry parcels with ghost buildings and the total number of parcels.	Program inception	ATD
<i>Registration Rate</i>	Registration Rate Percentage of ghost building parcels that get registered by the April 2011	2011	ATD
<i>Total expenditures</i>	Total local expenditures Per-resident	2000-2011	IMI Financial reports, <i>Quadro 3</i> SAIM
<i>Town Area Size</i>	Area Size of the town, in square km	2001	SAIM
<i>Altitude</i>	Altitude Altitude of the city, in meters	2001	SAIM
<i>Population</i>	Population Population, in thousand of inhabitants	2001	Census
<i>Disposable Income per capita</i>	Disposable income per capita at the municipal level, in thousand of euros	2005	SAIM
<i>Urbanization Index</i>	Index is equal to one if density is less than 100 people per sq. km; it is equal to two if density is between 100 and 500 people per sq. km; it is equal to three if density is above 500 people per sq.km.	2001	SAIM
<i>Justify Tax Cheating</i>	Answers to the question "Do you Justify Tax Cheating?", originally coded on a scale 1 (never justifiable) to 10 (always justifiable). Normalized variable.	1981-2008	EVS
<i>Speed of Public Good Provision</i>	Speed of current expenditures Ratio between paid over committed current expenditures	2005-2006	IMI

Notes: ATD stands for Agenzia del Territorio Database. IMI stands for Italian Ministry of the Interior. SAIM stands for Statistical Atlas of Italian Municipalities. EVS stands for European Values Survey.

Table B.2: *Political Variables' description and sources*

Variable	Definition and measure	Sample	Source
<i>Mayor Age</i>	Age of the mayor Age of the mayor, in number of years	Program inception	IMI
<i>Mayor Education</i>	Education of the mayor Categories: Primary Education, High school education, University degree, Postgraduate professional schooling, GED equivalent schooling, Vocational schooling	Program inception	IMI
<i>Mayor Born Same City</i>	Place of birth of the mayor Dummy variable equal to 1 if mayor is born in the same city	Program inception	IMI
<i>Mayor Term Number</i>	Tenure of the mayor Number of the mayoral's term, in number of years	Program inception	IMI
<i>Mayor Woman</i>	Gender of the mayor Dummy variable equal to 1 if the mayor is a woman	Program inception	IMI
<i>Incumbent Reelection</i>	Incumbent mayor is reelected Equal to 1 if mayor is reelected	1993-2011	IMI
<i>Incumbent Rerun</i>	Incumbent mayor decides to run for office again Equal to 1 if mayor re-runs	1993-2011	IMI
<i>N. Candidates</i>	Number of candidates Number of candidates	1993-2011	IMI
<i>Victory Margin</i>	Margin of victory Margin of victory of the winning candidate	1993-2011	IMI
<i>Runoff</i>	Election has a runoff Runoff	1993-2011	IMI

Notes: IMI stands for Italian Ministry of the Interior.

Appendix C

Appendix to Chapter 3

C.1 Extension to Model

Differentiating the social welfare function with respect to the length of leave, μ , we employ the envelope theorem and ignore both $\frac{\partial L_m}{\partial \mu}$ and $\frac{\partial L_w}{\partial \mu}$, as both men and women choose their labor supply optimally. We thus get:

$$\frac{\partial SW}{\partial \mu} = \frac{\partial W_m}{\partial \mu} \beta_m L_w + \left(\frac{\partial W_w}{\partial \mu} + v'(\mu) \right) L_w$$

Using men's and women's wage from expression 3.8, we obtain the following first order condition:

$$v'(\mu) = \left[\frac{\frac{\lambda+1}{\lambda} L_w - \frac{2-\lambda}{\lambda} \beta_m L_m}{2L_w - \beta_m L_m} \right] c'(\mu) \quad (\text{A1})$$

Note that when $\lambda = 1$ the first order conditions in expressions A1 and 3.9 are identical. This means that the weight on men's utility does not affect the analysis in the main text when $\lambda = 1$, but as will be now shown, it cannot be too large when $\lambda > 1$.

The second order condition that has to be met is:

$$\begin{aligned} v''(\mu) - \left[\frac{\frac{\lambda+1}{\lambda} L_w - \frac{2-\lambda}{\lambda} \beta_m L_m}{2L_w - \beta_m L_m} \right] c''(\mu) - \left[\frac{\left(\frac{\lambda+1}{\lambda} \frac{\partial L_w}{\partial \mu} - \frac{2-\lambda}{\lambda} \beta_m \frac{\partial L_m}{\partial \mu} \right) (2L_w - \beta_m L_m)}{(2L_w - \beta_m L_m)^2} \right] c'(\mu) \\ + \left[\frac{\left(2 \frac{\partial L_w}{\partial \mu} - \beta_m \frac{\partial L_m}{\partial \mu} \right) \left(\frac{\lambda+1}{\lambda} L_w - \frac{2-\lambda}{\lambda} \beta_m L_m \right)}{(2L_w - \beta_m L_m)^2} \right] c'(\mu) < 0 \end{aligned} \quad (\text{A2})$$

Since $c''(\mu) > 0 > v''(\mu)$, and since $\frac{\partial L_w}{\partial \lambda} > 0$ and $\frac{\partial L_m}{\partial \lambda} < 0$, the first three terms in expression A2 are negative, and only the fourth term is positive.

Assuming the second order condition holds, we can employ the implicit function theorem on the first order condition in expression A1 to obtain:

$$\begin{aligned} \text{sign}\left\{\frac{\partial \bar{\mu}}{\partial \lambda}\right\} = \text{sign}\left\{\left[\frac{1}{\lambda^2}(L_w - 2\beta_m L_m) - \frac{\lambda + 1}{\lambda} \frac{\partial L_w}{\partial \lambda} + \frac{2 - \lambda}{\lambda} \beta_m \frac{\partial L_m}{\partial \lambda}\right](2L_w - \beta_m L_m) \right. \\ \left. + (2 \frac{\partial L_w}{\partial \lambda} - \beta_m \frac{\partial L_m}{\partial \lambda}) \left(\frac{\lambda + 1}{\lambda} L_w - \frac{2 - \lambda}{\lambda} \beta_m L_m\right)\right\} \end{aligned} \quad (\text{A3})$$

Since $\frac{\partial L_w}{\partial \lambda} > 0$ and $\frac{\partial L_m}{\partial \lambda} < 0$, the second term in expression A3 is positive. The first term in expression A3 is positive as well, as long as β_m is not too large. Thus, as in the main text, we get that $\frac{\partial \bar{\mu}}{\partial \lambda} > 0$, that is the less society is willing to tolerate gender based discrimination the longer the maternity leave it will provide.

C.2 Data Appendix

- United Nations Statistical Division, Statistics and Indicators on Women and Men
 - Source: <http://unstats.un.org/unsd/demographic/products/indwm/statistics.htm>
 - Variables: Length of Maternity Leave (June 2010), Total Population (June 2010), Total Fertility Rate (June 2010), Percentage of Parliamentary Seats in Single or Lower Chamber Occupied by Women (2005).
- Penn World Table
 - Source: <http://pwt.econ.upenn.edu/>
 - Variables: Real Gross Domestic Product per Capita, current price (2006), Government Share of Real Gross Domestic Product per Capita, current price (2006).
- Ethnologue
 - Source: <http://www.ethnologue.com/web.asp>
 - Variable: Most widely spoken language in each country.
- CIA World Factbook
 - Source: <https://www.cia.gov/library/publications/the-world-factbook/fields/2098.html>
 - Variable: Most widely spoken language in each country.
- World Value Survey
 - Source: <http://www.worldvaluessurvey.org>

Table C.1: *Number of Cases of Gender Differentiated Pronouns*

Language	Gender Differentiated Personal Pronouns	Language	Gender Differentiated Personal Pronouns
Albanian	2	Italian	2
Arabic	4	Korean	2
Bulgarian	1	Latvian	2
Catalan	2	Lithuanian	2
Croatian	2	Luxembourgish	1
Czech	2	Macedonian	1
Danish	1	Mandarin	0
Dutch	1	Persian	0
English	1	Portuguese	2
Finnish	0	Romanian	2
French	2	Russian	1
German	1	Serbian	2
Greek	2	Spanish	4
Hebrew	4	Swedish	1
Hindi	0	Turkish	0
Hungarian	0	Ukrainian	1
Icelandic	2		

- Variables: Men should have more right to a job than women, Scale of income, Marital status, Age, Highest educational level attained.
- International Labor Organization, Maternity Protection Database
 - Source: <http://www.ilo.org/travaildatabase/servlet/maternityprotection>
 - Variable: Length of Paternity Leave

C.3 Language Coding

One of the stable grammatical features of a language is its use of personal pronouns. Personal pronouns are gender differentiated in some cases, but in other cases they are not. We code, using grammar books, the number of cases of gender differentiated personal pronouns in 33 languages. Table C.1 presents our data.